

14

50-203-

*Journal of*  
**Geophysical  
Research**

**VOLUME 64**

**JANUARY 1959**

**NUMBER 1**

**THE SCIENTIFIC PUBLICATION  
OF THE AMERICAN GEOPHYSICAL UNION**

# Journal of Geophysical Research

*An International Scientific Publication*

## OFFICERS OF THE UNION

MAURICE EWING, *President*  
LLOYD V. BERKNER, *Vice President*  
A. NELSON SAYRE, *General Secretary*  
WALDO E. SMITH, *Executive Secretary*

## OFFICERS OF THE SECTION

### *Geodesy*

MILTON O. SCHMIDT, *President*  
CHARLES PIRCE, *Vice President*  
FRANK L. CULLEY, *Secretary*

### *Siesmology*

HUGO BENIOFF, *President*  
LEONARD M. MURPHY, *Vice President*  
JAMES A. PEOPLES, JR., *Secretary*

### *Meteorology*

HELMUT E. LANDSBERG, *President*  
THOMAS F. MALONE, *Vice President*  
WOODROW C. JACOBS, *Secretary*

### *Geomagnetism and Aeronomy*

H. R. JOSETING, *President*  
L. R. ALLDREDGE, *Vice President*  
ROBERT E. GEBHARDT, *Secretary*

### *Oceanography*

ROGER R. REVELLE, *President*  
HENRY STOMMEL, *Vice President*  
DONALD W. PRITCHARD, *Secretary*

### *Volcanology, Geochemistry, and Petrology*

J. FRANK SCHAIRER, *President*  
FRANCIS G. WELLS, *Vice President*  
L. T. ALDRICH, *Secretary*

### *Hydrology*

RAY K. LINSLEY, *President*  
HARRY F. BLANEY, *Vice President*  
RALPH N. WILSON, *Secretary*

### *Tectonophysics*

HARRY H. HESS, *President*  
PATRICK M. HURLEY, *Vice President*  
BENJAMIN F. HOWELL, JR., *Secretary*

## BOARD OF EDITORS

Editors: PHILIP H. ABELSON and J. A. PEOPLES, JR.

## ASSOCIATE EDITORS

1959

JULIUS BARTELS	D. F. MARTYN
JOHN W. EVANS	TOR J. NORDENSON
H. W. FAIRBAIRN	HUGH ODISHAW
JOSEPH KAPLAN	E. H. VESTINE
THOMAS MADDOCK, JR.	J. LAMAR WORZEL

1959-1960

HENRY G. BOOKER	WALTER B. LANGBEIN
E. C. BULLARD	ERWIN SCHMID
JULE CHARNEY	HENRY STOMMEL
GEORGE T. FAUST	J. TH. THIJSSSE
DAVID G. KNAPP	A. H. WAYNICK

J. TUZO WILSON

1959-1961

HENRI BADER	T. NAGATA
K. E. BULLEN	FRANK PRESS
CONRAD P. MOOK	A. NELSON SAYRE
WALTER H. MUNK	MERLE A. TUVE
JAMES A. VAN ALLEN	

This Journal welcomes original scientific contributions on the physics of the earth and its environment.

Manuscripts should be transmitted to J. A. Peoples, Jr., Geology Department, University of Kansas, Lawrence, Kansas. Authors' institutions, if in the United States or Canada, are requested to pay a publication charge of \$15 per page, which, if honored, entitles them to 100 free reprints.

Subscriptions to the *Journal of Geophysical Research* and *Transactions, AGU* are included in membership dues.

Non-member subscriptions, *Journal of Geophysical Research*. . . . . \$16 per calendar year, \$2 per copy

Non-member subscriptions, *Transactions, AGU*. . . . .

. . . . . \$4 per calendar year, \$1.25 per copy

Subscriptions, renewals, and orders for back numbers should be addressed to American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington 5, D. C. Suggestions to authors are available on request.

Advertising Representative: Howland and Howland, Inc., 114 East 32nd St., New York 16, N. Y.

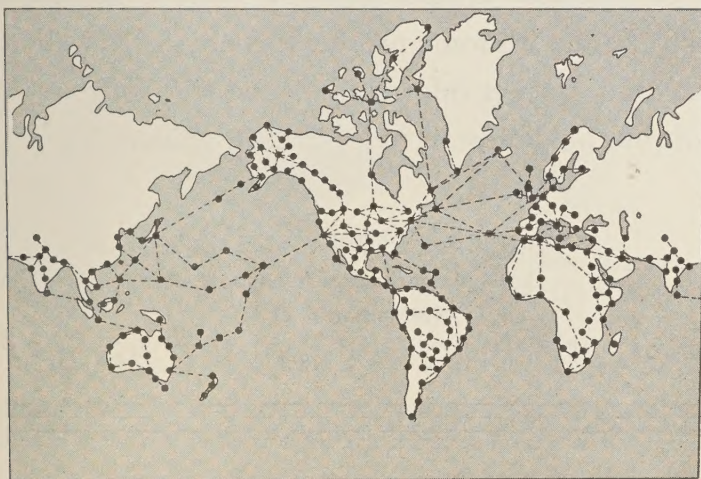
Beginning with the January 1959 issue (Vol. 64, No. 1) the *Journal of Geophysical Research* is published monthly by the American Geophysical Union, 1515 Massachusetts Ave., Northwest, Washington 5, D. C., with the support of the Carnegie Institution of Washington and the National Science Foundation. The new monthly combines the type of scientific material formerly published in the bi-monthly *Transactions, American Geophysical Union*, and the quarterly *Journal of Geophysical Research*. The *Transactions, American Geophysical Union* will continue as a quarterly publication for Union business and items of interest to members of the Union.

Second-class postage paid at Richmond, Virginia.



112714

# GEODETIC GRAVITY METER



**WORLD SURVEY WITHOUT RESETTING**

**Negligible Drift**

**LESS THAN .5 mgl PER MONTH**

**Stable Calibration**

**NEVER REQUIRES RECALIBRATING**

**No "Sets" or "Tares"**

**UNDER NORMAL OPERATION**

*The Ultimate in Accuracy and Reliability in*

**LAND, UNDERWATER, GEODETIC & EARTH-TIDE GRAVITY METERS**

**LaCoste & Romberg**

*Gravity Meters —*

6606 NORTH LAMAR . . . . .

AUSTIN, TEXAS

## RUSSIAN TRANSLATORS WANTED

In connection with the production of an English edition of *Izvestiia* (Geophysics Series) of the Academy of Sciences, USSR, and possibly other Russian scientific publications, additional translators are needed. It is essential that translators be conversant with some phase of geophysics as well as with Russian. For further information, send word of your interest and of your field of proficiency in geophysics to

**AMERICAN GEOPHYSICAL UNION**  
1515 Massachusetts Avenue, N. W.  
Washington 5, D. C.

READ the new monthly

## JOURNAL OF GEOPHYSICAL RESEARCH

This monthly contains the scientific and technological material formerly carried in the bimonthly *Transactions*, *American Geophysical Union*, and in the quarterly *Journal of Geophysical Research*. In addition, it carries results of the scientific findings of the IGY.

Read the  
Journal of Geophysical Research

Advertise in its pages  
Patronize its Advertisers  
Subscriptions

\$16.00 per calendar year  
\$2.00 per copy

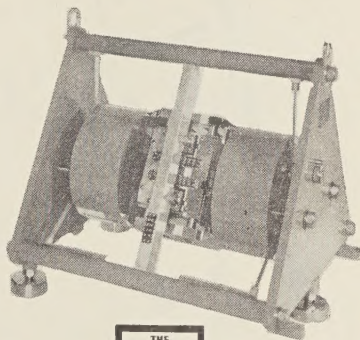
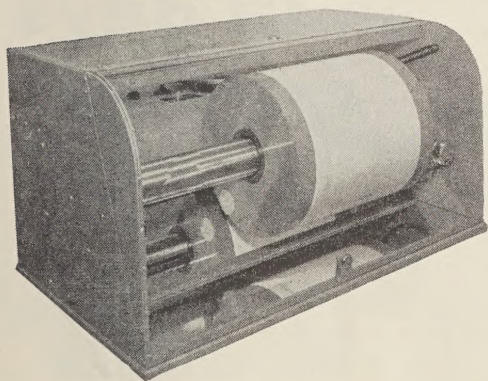
American Geophysical Union

1515 Massachusetts Avenue, N. W.

Washington 5, D. C.



*Manufacturers of*  
**Seismological Station  
 Instruments**



**Inquiries Invited  
 WRITE TO**

**The GEOTECHNICAL CORPORATION**

**GEOPHYSICAL MONOGRAPH SERIES**

**Antarctica in the International Geophysical Year**—Geophysical Monograph No. 1 (Publication No. 462, National Academy of Sciences—National Research Council), 133 pp. and large folded map of the Antarctic, 1956, 7" x 10" **\$6.00**

Contains 16 separate papers by various American authorities on the Antarctic under the headings: General, Geographic and Meteorological, Geological and Structural, Upper Atmospheric Physics, and Flora and Fauna. Map (41" x 41") compiled by the American Geographical Society. Introduction by L. M. Gould, President of Carleton College and internationally recognized authority on the Antarctic.

**Geophysics and the IGY**—Geophysical Monograph No. 2 (Publication No. 590, National Academy of Sciences—National Research Council), 210 pp., 1958, 7" x 10" **\$8.00**

Contains 30 separate papers by leading American authorities under the headings: Upper Atmospheric Physics, The Lower Atmosphere and the Earth, and The Polar Regions. Preface by Joseph Kaplan, Chairman of the U. S. National Committee for the IGY.

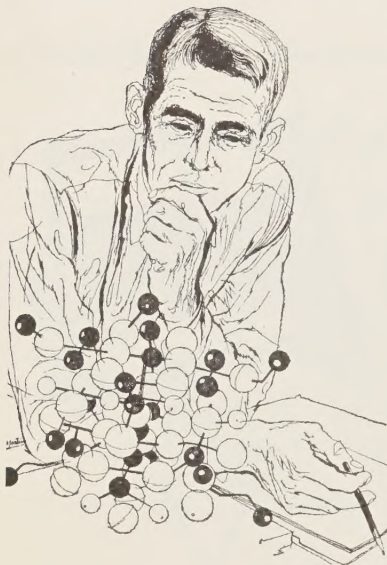
*Prices plus postage, unless payment accompanies order. Quantity discounts: 5–19 copies, 10%; 20–49 copies, 15%; 50 or more copies, 20%.*

**AMERICAN GEOPHYSICAL UNION**

1515 MASSACHUSETTS AVENUE, N.W.

WASHINGTON 5, D. C., U.S.A.

Please mention JOURNAL OF GEOPHYSICAL RESEARCH, when writing to advertisers



## CREATIVE ENGINEERING at HUGHES

Freedom of investigation...unmatched laboratory facilities...financial support of efforts toward advanced degrees...these are the pillars of what the Hughes Research & Development Laboratories refer to as creative engineering. These pillars, combined with the diversified Hughes activity in almost all fields, result in a highly favorable environment for the engineer or physicist capable of making significant contributions in his field of specialization. Opportunities for creative engineers at Hughes have never been more promising! Investigate by writing directly to:

Dr. Allen Puckett  
Associate Director  
Systems Development Laboratories

*The West's leader in advanced electronics*

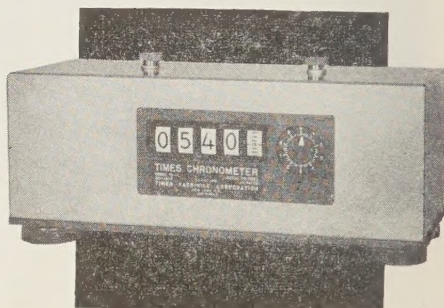
**HUGHES**

Hughes Aircraft Company

Culver City 6, California

# THE TIME INDICATOR UNIT

## accurate to 1 second in 12 days



### TIMES MODEL TS-3 CHRONOMETER

Program timer, pulse generator and clock. Timing assemblies, driven by the clock motor, provide momentary contact closings at rate of

- ONCE A SECOND
- ONCE A MINUTE
- ONCE AN HOUR

also optional frequency or pulse outputs as specified in range between 10 and 1000 cps.

PRICE: \$950.00, F.O.B. Factory.

Optional frequency output, \$50.00 each.

*Write for details.*



# TIMES FACSIMILE

CORPORATION

540 W. 58th St. New York 19, N. Y.



# Journal of GEOPHYSICAL RESEARCH

VOLUME 64

JANUARY, 1959

No. 1

## Densities and Temperatures of the Upper Atmosphere Inferred from Satellite Observations

G. F. SCHILLING AND T. E. STERNE

*Smithsonian Astrophysical and Harvard College Observatories,  
Cambridge, 38, Massachusetts*

**Abstract**—Surface observations of artificial Earth satellites permit the inference of upper atmosphere densities from changes in the orbital elements caused by drag. The atmospheric density between some 180 km to 400 km altitude appears to be appreciably higher, on the average, than was usually thought before satellite data became available. In order to be consistent with the observed phenomena, the atmospheric temperature at or below these altitudes must also be higher than that given by a number of model atmospheres.

**Introduction**—Since the launching of the first artificial Earth satellite in October, 1957, sufficient information has become available for the derivation of values of upper atmosphere density at various altitudes. We have collected such studies as conducted by different investigators, and discuss here the results and implications of all similar investigations known to us at this time. The conclusions which we are able to draw reflect the potential importance of this new tool of research to our understanding of the physics and chemistry of the upper atmosphere.

**Observational values**—Table 1 contains a full tabulation of numerical values of these atmospheric densities, arranged according to geometric height above the Earth's surface. The column headed Satellite refers to the object which served as the observational tool for each author.

It will be of value to define the scientific terminology commonly used [Whipple and others, 1958] for the artificial satellites mentioned in this discussion. Table 2 lists the satellites and carrier rockets launched successfully into orbit to date (June, 1958). The initial values of apogee and perigee heights are given in kilometers above the geoid for the date of launching. While the perigee distance from the Earth's center decreases in time, the actual perigee height above

TABLE 1—Atmospheric densities derived by various investigators

Height	Density	Satellite	Reference
km	gm/cm <sup>3</sup>		
656	$3.5 \times 10^{-18}$	1958 $\beta 2$	Jacchia [1958b]
368	$1.5 \times 10^{-14a}$	1958 $\alpha$	Sterne [1958a]
368	$1.4 \times 10^{-14}$	1958 $\alpha$	Sterne [1958a]
275	$8.5 \times 10^{-14}(?)$	1957 $\alpha 2$	Harris and Jastrow [1958] <sup>b</sup>
241	$2.5 \times 10^{-13}$	1957 $\alpha 2$	Royal Aircraft [1957]
233	$2.2 \times 10^{-13}$	1957 $\beta 1$	Sterne and Schilling [1958]
232 $\pm$ 5	$1.5 \times 10^{-13}(?)$	1957 $\alpha 2$	Harris and Jastrow [1958] <sup>b</sup>
220	$5.7 \times 10^{-13}$	1957 $\alpha 1$	Sterne and Schilling [1958]
220	$4.5 \times 10^{-13a}$	1957 $\alpha 2$	Sterne and Schilling [1958]
220	$4.0 \times 10^{-13}$	1957 $\alpha 2$	Sterne [1958a]
220	$4.0 \times 10^{-13}$	1957 $\alpha 2$	Warwick [1958]
215	$4.7 \times 10^{-13}$	1957 $\alpha 2$	Priester and others [1958]
212	$4.8 \times 10^{-13}$	1957 $\beta 1$	Sterne and Schilling <sup>c</sup>
212	$4.4 \times 10^{-13a}$	1957 $\beta 1$	Sterne and Schilling <sup>c</sup>
211 $\pm$ 4	$4.6 \times 10^{-13}$	1957 $\beta$	Groves [1958]
206 $\pm$ 7	$5.4 \times 10^{-13}$	1957 $\alpha 2$	Groves [1958]
202 $\pm$ 4	$7.3 \times 10^{-13}$	1957 $\alpha 1$	Groves [1958]
201 $\pm$ 4	$6.7 \times 10^{-13}$	1957 $\alpha 2$	Groves [1958]
200 <sup>e</sup>	$4.0 \times 10^{-13}$	1957 $\alpha 2$	Mullard Observatory [1957]
197 $\pm$ 1	$7.0 \times 10^{-13}$	1957 $\beta$	Groves [1958]
186	$6.7 \times 10^{-13}$	1958 $\gamma$	Sterne [1958b]

<sup>a</sup> Result of early calculations which have been superseded by later studies, based on the same orbital information.

<sup>b</sup> Harris and Jastrow have published three suggested extrapolated variations of density with altitude. The value of  $1.5 \times 10^{-13}$  gm/cm<sup>3</sup> appears to be the actual mean density value derived by them from Minitrack observations for the perigee altitude of 232  $\pm$  5 km at the epoch given.

<sup>c</sup> Unpublished.

TABLE 2—Characteristics and terminology of satellites

Satellite	Name	Date of launching	Initial apogee height	Initial perigee height	Initial period
		GMT	km	km	min
1957 α1	Sputnik I carrier rocket	Oct. 4, 1957	950	225	96.2
1957 α2	Sputnik I	Oct. 4, 1957	950	225	96.2
1957 β1	Sputnik II	Nov. 3, 1957	1670	240	103.7
1958 α	Explorer I	Feb. 1, 1958	2540	368	114.95
1958 β1	Vanguard carrier rocket	Mar. 17, 1958	3965	652	134.29
1958 β2	Vanguard	Mar. 17, 1958	3965	652	134.29
1958 γ	Explorer III	Mar. 26, 1958	2800	188	115.91
1958 δ1	Sputnik III carrier rocket	May 15, 1958	1880	241	105.9
1958 δ2	Sputnik III	May 15, 1958	1880	241	105.9

the Earth's surface varies and can even slightly increase above its initial value in accord with the latitudinal position of the perigee point over an oblate Earth.

Figure 1 shows a plot of density versus altitude on a semi-logarithmic scale. In addition to most of the observational values of Table 1, the figure contains portions of model atmospheres between 175 km and 400 km.

It is at once apparent that all individual density values fall within an area bounded by the RAND Model I Atmosphere for 45° latitude [Grimminger, 1948] and the ARDC Model

Atmosphere [Minzner and Ripley, 1956]. Within this area, Models 1 and 2 of the Smithsonian Interim Atmosphere [Sterne and others, 1957] appear to approximate the observations rather closely. Although the three Smithsonian interim models were based essentially on only one reliable early density, inferred from Satellite 1957 α2, each was internally consistent.

The Rocket Panel [1952] Atmosphere reaches to 220 km. The Naval Research Laboratory rocket observation, made on August 7, 1951, and used as the basis for the Model at this high altitude, gave a value of  $1.0 \times 10^{-13}$  g/cm<sup>3</sup> at 219 km.

*Discussion*—We want to add a few critical comments of a specific nature. The density values derived from the USSR 1957 α1 and 1957 β1 satellites must be considered as being of low reliability, since, to our knowledge, the precise technical rocket parameters have not yet been announced. Reasonable parameters of the mass-area ratios have been derived [Groves, 1958; Schilling and Rinehart, 1958] only as speculative estimates. With regard to Satellite 1957 α2, we do not know to what extent particular authors have taken into account the effects of aerodynamic drag on the antennas. We have found [Sterne and Schilling, 1958] that antenna drag decreases the effective mass-area

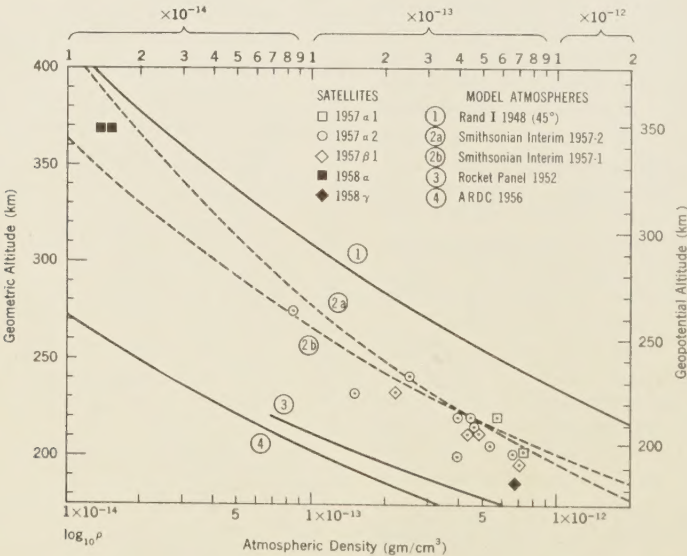


FIG. 1—Variation of atmospheric density with altitude



ratio for this spherical satellite by about 25 per.

In all of our own determinations of density from satellites, we have applied the assumption that a representative sample of all possible orientations has occurred at perigee. We know that different satellites have undergone variable accelerations, and it has actually been possible to deduce aspect variations from fluctuations in the drag effect [Whitney, 1958a]. But all density points determined by us must be considered as mean values, averaged over a large number of consecutive revolutions. If and when the assumption of random orientation does not apply, density determinations should therefore spread to both higher and lower values from the mean points plotted in Figure 1. So far, we have been able to study this effect satisfactorily only in the case of Satellite 1957  $\beta$ .

For a non-spherical object, we are interested in the effective presentation area during its flight through perigee. For this purpose we have used an effective mass-area ratio  $m/A$ , where  $m$  is the mass of the object, and  $A$  is the average cross-sectional area of the object projected on a plane normal to the direction of flight. If all orientations of the object occur with equal frequency relative to the direction of flight,  $A$  is the total superficial area divided by 4. It can be proved generally that the value of  $A$ , averaged over all orientations of any convex body, is exactly  $\frac{1}{4}$  of the total superficial area.

In addition, it must be remembered that the arguments of perigee varied with time as a function of the orbital inclinations. For the Soviet satellites this amounted to a daily shift in geographical latitude of the order of  $0.4^\circ/\text{day}$ ; for the USA satellites the change was sometimes as much as about  $3.5^\circ/\text{day}$ . In consequence, all density values collected in Table 1 should refer to a specific geographical latitude, depending on the epoch or time of observation. A recent discussion of this phenomenon has been published by Groves [1958].

However, we have critically examined the practicability of making such distinctions and conclude that such detailed considerations would be misleading on the basis of presently available observational material. The variability of the attitude and, hence, possible uncertainties in the effective presentation area of all satellites discussed here, is still numerically larger than any density variation with latitude or season that

may reasonably be expected. For a striking example, refer to studies by Jacchia [1958a] and Whitney [1958b] on the variable acceleration of the satellites. It appears mandatory to us, then, to treat all values obtained to date as average densities for a given height above the geoid.

*Conclusions*—All values of density listed in Table 1, derived from the changes in the orbital elements of artificial Earth satellites, relate to altitudes near or above the respective perigees at various geographical latitudes and are yet too few to scatter about an unbiased mean. Since a number of simplifying assumptions are implicitly involved in any method developed to date, the absolute reliability of the numerical results is considerably reduced for non-spherical satellites. Because of antenna drag effects, this category includes all objects which have been launched into satellite orbits so far. There is also a possibility that the apparently higher densities may be due in part to additional drag forces of electrostatic origin [Jastrow and Pearse, 1957]. A critical study of this possibility [Sterne, 1958b] has, however, placed substantial limits on the reality of such an effect at the altitudes of present concern. We can thus conclude that artificial satellites may soon constitute the most effective tool for accurately measuring such variations of density of the atmosphere over a wide range of latitude, altitude, and time.

It is of particular importance to explore the implications which these new findings of density carry with regard to the distribution of temperature in the upper atmosphere. We have already shown a possible way to account for the higher densities in terms of higher atmospheric temperatures and pressures [Sterne and others, 1957]. In devising interim atmospheric models to agree with the densities inferred from Satellite 1957  $\alpha 2$  at an altitude of 220 km, we were forced to adopt considerably higher temperatures below this level than the values adopted in the ARDC Model [Minzner and Ripley, 1956]. We are aware that our interim model atmosphere is crude and only roughly approximate, and of use principally for predicting orbital accelerations and lifetimes of new Earth satellites. However, we do not think the time has quite come to change our interim atmosphere in the light of the additional density values which have since been inferred from satellites (see Fig. 1).

If we were to propose a more realistic model for

the distribution of the molecular scale temperature with height on the basis of the now-available density data from satellites, we would feel inclined to consider seriously the possibility that a close correlation exists between temperature and the variation of electron and ion density with altitude. In other words, instead of assuming a constant temperature gradient, we would investigate the plausibility of humps of higher temperature in the  $F_1$ ,  $F_2$ , and  $G$  regions of the ionosphere. However, since satellites, and especially spherical satellites, will soon provide quantities of reliable new observational data, it is superfluous for us to extend such speculations further at this time.

Obviously, these newest results with regard to atmospheric densities inferred from satellite observations seem to support, to a certain degree, the pioneering study made by *Grimminger* [1948] which led to the early RAND Atmospheric Models. His numerical results at these altitudes were based to a great extent on extrapolations of results derived from one of the oldest tools of upper-atmosphere research, namely meteor observations [Whipple, 1951].

Finally, all densities furnished so far by the artificial satellites appear to lie between those atmospheric models derived from rocket data and those derived from meteor observations.

#### REFERENCES

- GRIMMINGER, G., Analysis of temperature, pressure, and density of the atmosphere to extreme altitudes, *RAND Report R-105*, Rand Corporation, Santa Monica, Calif., 149 pp., 1948.
- GROVES, G. V., Effect of the Earth's equatorial bulge on the life-time of artificial satellites and its use in determining atmospheric scale-heights, *Nature*, **181**, 1055, April 12, 1958.
- HARRIS, I., AND R. JASTROW, Upper-atmosphere densities from Minitrack observations on Sputnik I, *Science*, **127**, 471-472, February 28, 1958.
- JACCHIA, L. G., Basic orbital data for Satellite 1957 Beta One, *Smithson. Contr. Astrophys.*, **2**, 285-286, 1958a.
- JACCHIA, L. G., The secular perturbations and the orbital acceleration of Satellite 1958 Beta Two, *Smithson. Astrophys. Obs., Spec. Rep. 12*, pp. 30-33, 1958b.
- JASTROW, R., AND C. A. PEARSE, Drag on the satellite, *J. Geophys. Res.*, **62**, 413-423, 1957.
- MINZNER, R. A., AND W. S. RIPLEY, The ARDC model atmosphere, 1956, *Air Force Surveys in Geophysics 86*, 201 pp., Geophysics Research Directorate, AFRCRC, ARDC, December 1956.
- MULLARD RADIO ASTRONOMY OBSERVATORY, Radio observations of the orbit of the first Russian Earth satellite, *Nature*, **180**, 879-883, November 2, 1957.
- PRIESTER, W., H. G. BENNEWITZ, AND P. LENGUESSER, *Radiobeobachtungen des ersten kuenstlichen Erdsatelliten*, Westdeutscher Verlag, Cologne and Opladen, 38 pp., 1958.
- ROCKET PANEL, Pressures, densities, and temperature in the upper atmosphere, *Phys. Rev.*, **88**, 1027-1032, 1952.
- ROYAL AIRCRAFT ESTABLISHMENT, Observations of the orbit of the first Russian Earth satellite, *Nature*, **180**, 937-941, November 9, 1957.
- SCHILLING, G. F., AND J. S. RINEHART, Note on the mass-area ratios of the USSR satellites, *Smithson. Astrophys. Obs., Spec. Rep. 12*, pp. 20-23, 1958.
- STERNE, T. E., The density of the upper atmosphere, *Smithson. Astrophys. Obs., Spec. Rep. 11*, pp. 18-22, 1958a.
- STERNE, T. E., High-altitude atmospheric density, *Physics of Fluids*, **1**, 165-170, 1958b.
- STERNE, T. E., B. M. FOLKART, AND G. F. SCHILLING, An interim model atmosphere fitted to preliminary densities inferred from USSR satellites, *Smithson. Astrophys. Obs., Spec. Rep. 7*, 8 pp., 1957 (reprinted in *Smithson. Contr. Astrophys.*, **2**, 275-279, 1958).
- STERNE, T. E., AND G. F. SCHILLING, Some preliminary values of upper atmosphere density from observations of USSR satellites, *Smithson. Contr. Astrophys.*, **2**, 207-210, 1958.
- WARWICK, J. S., Radio observations of USSR satellites 1957  $\alpha 2$  and 1957  $\beta$ , *Sci. Rep. 10, High Altitude Obs., Univ. of Colo.*, ARDC Contract AF 19(604)-1491, 52 pp., April 30, 1958.
- WHIPPLE, F. L., Meteors as probes of the upper atmosphere, *Compendium of Meteorology*, (T. F. Malone, ed.), pp. 356-365, Amer. Met. Soc., 1951.
- WHIPPLE, F. L., L. G. BOYD, J. A. HYNEK, AND G. F. SCHILLING (ed.), Orbital data and preliminary analyses of satellites 1957 Alpha and 1957 Beta, *Smithson. Contr. Astrophys.*, **2**, 189-347, 1958.
- WHITNEY, C. A., The orbit and variable acceleration of satellite 1958 Alpha, *Smithson. Astrophys. Obs. Spec. Rep. 11*, pp. 14-17, 1958a.
- WHITNEY, C. A., The acceleration of Satellites 1958 Alpha and 1958 Gamma, *Smithson. Astrophys. Obs., Spec. Rep. 12*, p. 29, 1958b.

(Manuscript received June 27, 1958; presented at the Thirty-Ninth Annual Meeting, Symposium on Satellite Geophysical Studies, American Geophysical Union, Washington, D. C., May 6, 1958.)



## The Diurnal and Annual Variations of $foF2$ over the Polar Regions

S. C. CORONITI AND R. PENNDORF

*Electronics Research Laboratory  
Research and Advanced Development Division  
AVCO Manufacturing Corporation  
Boston, Massachusetts*

**Abstract**—Critical frequency data,  $foF2$ , are analyzed in the polar regions for the period 1954 to summer 1957. The diurnal variation is largest in winter and smallest in summer. In winter little or no diurnal variation occurs north of  $75^\circ$ . The amplitude increases southward up to 11 Mc/s at  $60^\circ\text{N}$  with a minimum around 07h and a maximum between 12h to 14h. The daily variation is fairly symmetrical around the maximum. The pronounced daily variation lasts from fall through winter till spring. The annual variation shows a regular behavior, with steep gradients occurring around sunrise and sunset throughout the year. The latitudinal differences are small in summer, of the order of 1 Mc/s or less between  $60^\circ$  and  $80^\circ\text{N}$ , but large in winter. The lines of equal  $foF2$  seem to fall between the geographic and geomagnetic latitude circles. The diurnal variation in the Antarctic is different from that in the Arctic during the southern summer (November to February) because a dip occurs around noon.

**Introduction**—Surveying our present knowledge of the  $F$  region, Martyn [1955a, p. 212] states that ionization production in the  $F$  region is caused solely by the sun. The total ionization density taken over the vertical extent of the whole  $F$  region seems to follow the zenith angle variation of the sun, as shown by Ratcliffe [1951], and Chatterjee [1953], although Skinner and Wright [1954] do not confirm Chatterjee's results. The maximum electron concentration of the  $F2$ -layer,  $foF2$ , however, does not follow the zenith angle of the sun, as is the case for the  $E$ - and  $F1$ -layers. As is well known, the morphology of  $foF2$  is definitely unusual in space and time. A geomagnetic influence on the  $F2$ -layer has been established in low and middle latitudes by Martyn [1955b, p. 260].

Allen [1953] conducted a thorough analysis of the diurnal variations of  $F2$  critical frequencies and virtual heights versus the phase of the sunspot cycle, season, geographic and geomagnetic latitudes. Unfortunately, his analysis of ten stations does not include Arctic stations, but it contains a complete description of the behavior of  $foF2$  in middle and low latitudes. We can, therefore, relate our results to the established facts in middle latitudes.

Since no over-all study of Arctic stations exists, we have analyzed all of the available  $foF2$  ionospheric data published in the CRPL series.

The analysis was restricted to that data obtained at stations located above  $60^\circ\text{N}$  and below  $60^\circ\text{S}$ . The data from seventeen stations were evaluated for the period of 1954 to 1957. The restriction of the number of stations was due to large gaps in the data at some stations and to the malfunction of equipment at others. Despite the fact that their operations started in 1956, Thule and Longyearbyen are included because they represent the most northerly stations on the globe.

The monthly median  $foF2$  values have been plotted from January 1954 to December 1957. Figures 1 (a, b, c, and d) show the four months selected from this material. All of the available data, for three or four years, are presented for each station; the abscissa is mean local time; the ordinate is the value of  $foF2$  in Mc/s. Shown on these curves are two arrows indicating sunrise ( $\uparrow$ ) and sunset ( $\downarrow$ ) at an altitude of 300 km. The same data are presented in time maps in Figures 2 (a, b, and c) which show the daily, the annual, and the solar activity variations. The months starting with January 1954 and ending in summer 1957 are shown along the abscissa, and the local mean time starting at midnight is shown along the ordinate. Isopleths are drawn for fixed critical frequencies in steps of 1 Mc/s and in steps of 2.5 Mc/s for  $foF2 \geq 7.5$  Mc/s. Shading indicates the time variation for the selected frequency interval.

*Daily variation in summer*—The daily variation can be observed from a study of Figures 1 or 2. It is greatest during the polar winter. The summer curves (for the months of May, June, July, and August) and the time maps for all

stations are characterized by a high degree of constancy in the critical frequencies. No marked diurnal variation is visible. This is not surprising because the sun never sets at  $F_2$ -layer height over the polar regions.

## MARCH

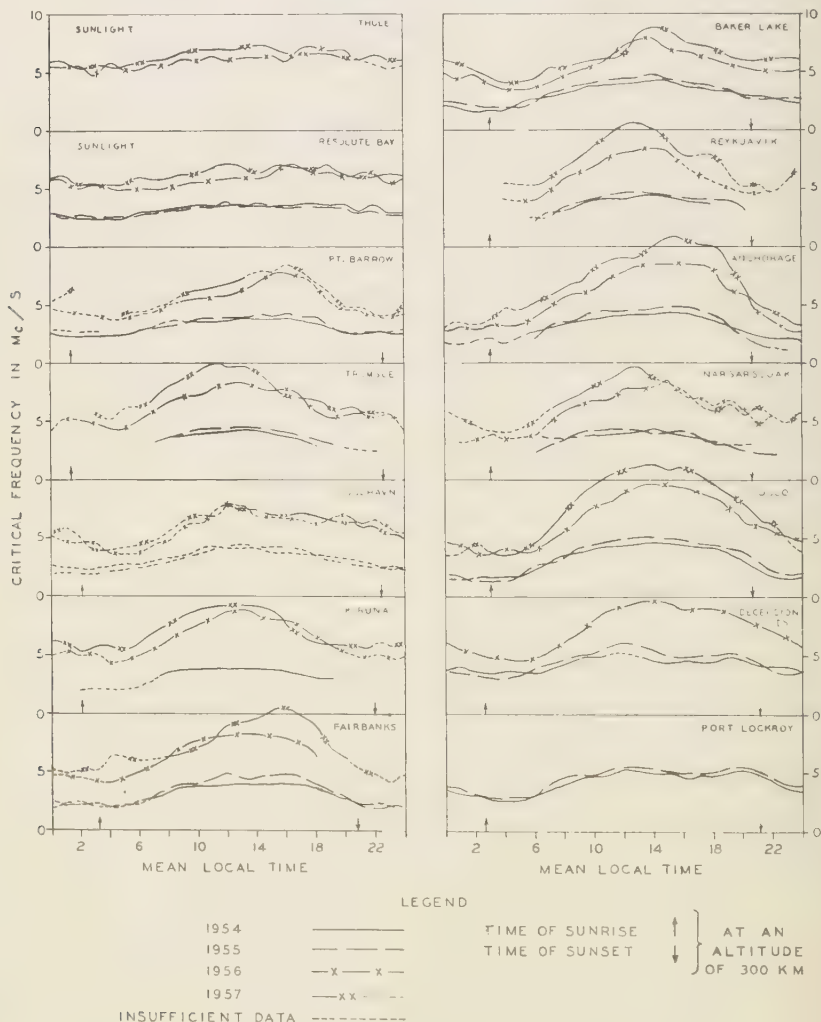


FIGURE 1A

Daily variation of the monthly median values of the critical frequency,  $foF_2$ , for latitudes  $60-90^\circ$  N and S.

Four months (March, June, September, December) are selected.

Hourly values are given for the period 1954 to summer 1957.



In July 1956, the average monthly median value for the three stations, Thule, Resolute Bay, and Godhavn, is 5.4 Mc/s; for the stations beneath the auroral zone it is 6.0 Mc/s, and for the more southerly stations it is 5.7 Mc/s. The

maximum diurnal range in critical frequencies for Thule and Resolute Bay is only 0.8 Mc/s, but is 1.7 Mc/s for Godhavn. Under the auroral zone it increases to 2.5 Mc/s and reaches 2.7 Mc/s farther south. Thus the diurnal range

## JUNE

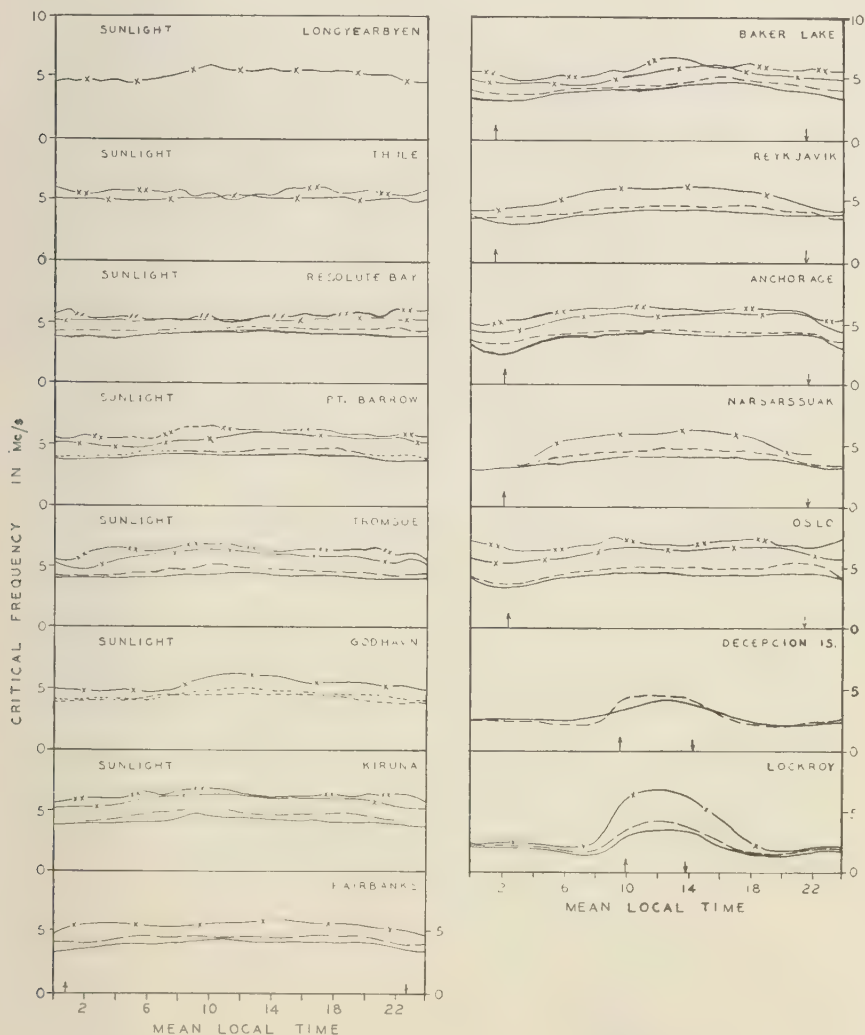


FIGURE 1B

Daily variation of the monthly median values of the critical frequency, foF2, for latitudes 60°-90° N and S.

Four months (March, June, September, December) are selected.

Hourly values are given for the period 1954 to summer 1957.



mated under these conditions. At Longyearbyen and Thule, admitting a deduction based on meager data, this equilibrium persists even during the winter months. This fact is very significant. It indicates that since the rate of production and the rate of destruction of elec-

trons are nearly equal, the mechanism must be one in which the processes are continuous. Considering present uncertainties in reading and in scaling of the records, it is not certain that the difference between auroral zone and the more southerly stations is significant. During June

## DECEMBER

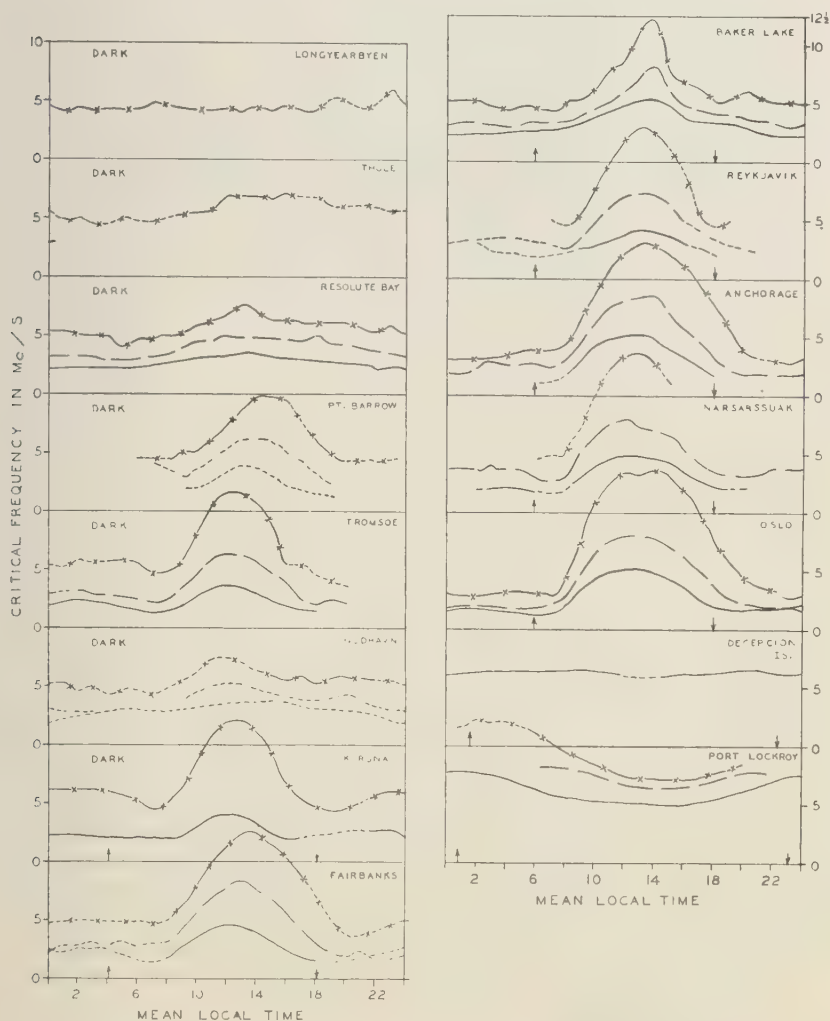


FIGURE 10

Daily variation of the monthly median values of the critical frequency, foF2, for latitudes 60°-90° N and S.

Four months (March, June, September, December) are selected.

Hourly values are given for the period 1954 to summer 1957.



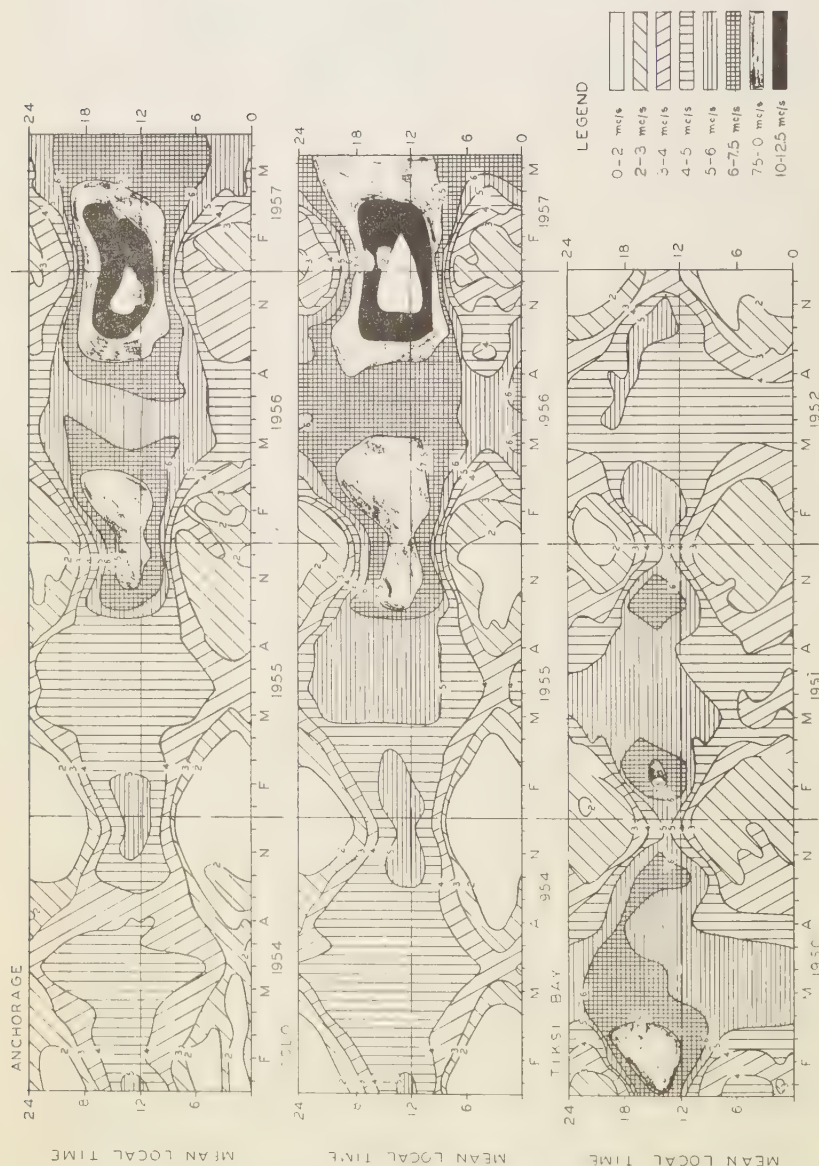


FIGURE 2A

Isoleths for the critical frequency of the F2-layer,  $foF_2$  in mc/s. Selected stations north of  $60^\circ$ . Data are given for the period 1952 to summer 1957, except for Tiksi, USSR, where the period 1950 to 1952 is selected.

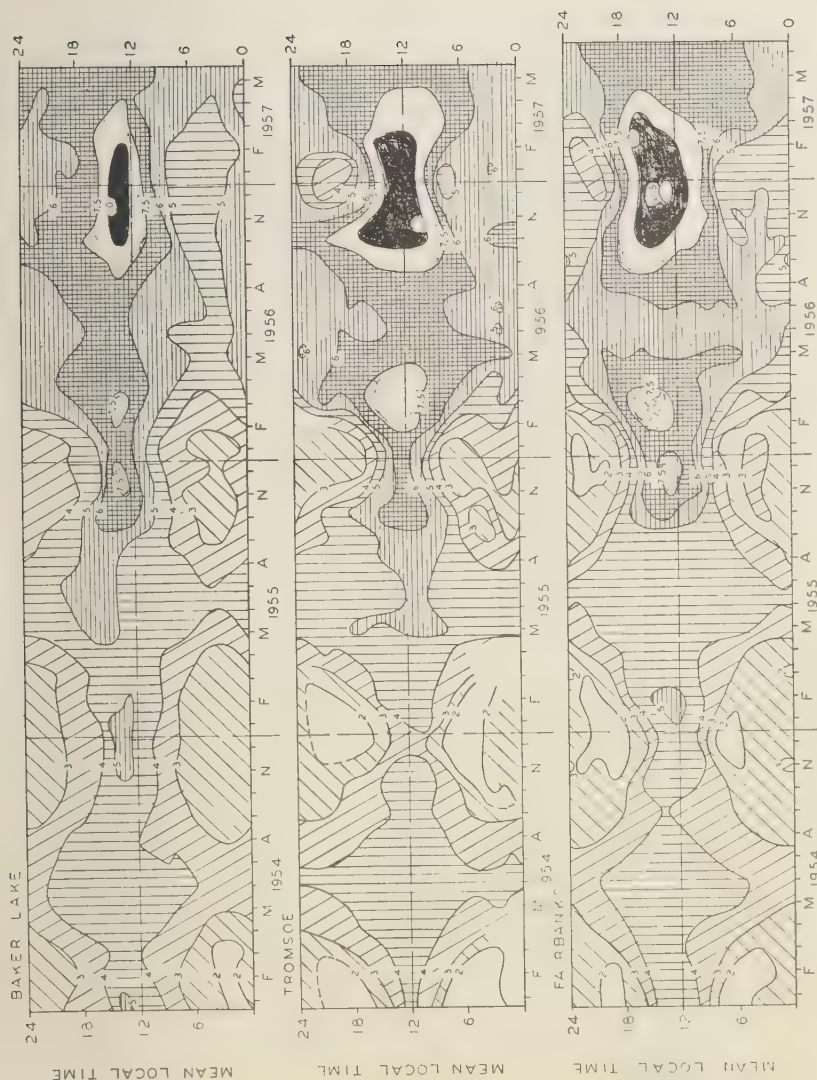
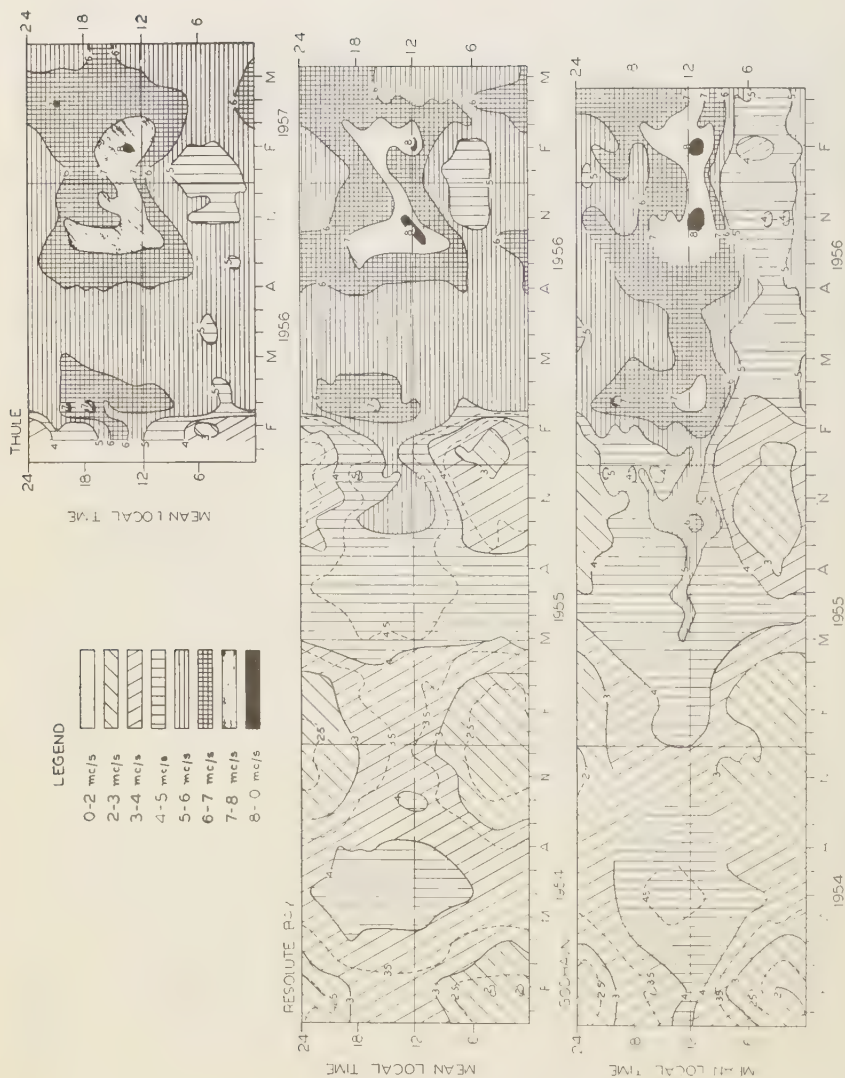


FIGURE 28  
Isoleths for the critical frequency of the F2-layer, foF2 in mc/s.  
Selected stations north of 60°. Data are given for the period  
1952 to summer 1957, except for Tiksi, USSR, where the period  
1950 to 1952 is selected.



**FIGURE 2C**  
 Isopleths for the critical frequency of the F<sub>2</sub>-layer, for 60°N.  
 Selected stations north of 60°. Data are given for the period  
 1952 to summer 1957, except for Tiksi, USSR, where the period  
 1950 to 1952 is selected.



and July the sun never sets at  $F2$ -layer heights; it stays within a limited range of elevation angle. Some slight daily variation, however, seems to begin to appear for stations south of  $64^\circ\text{N}$ . The more or less constant June data indicate clearly that the altitude of the sun has no appreciable effect on  $f_oF2$ ; that is, the Chapman theory is not applicable.

The behavior during the southern summer, however, deviates markedly from that during the northern summer. The December data for Deception Island and Port Lockroy show high values at night and a minimum during the day. Furthermore, the amplitude of the daily variation is very large in Port Lockroy, 2.7 Mc/s in December 1954 and 4.8 Mc/s in December 1956. Such values are far in excess of those cited above for the northern hemisphere. This strange daily variation with a minimum between 12h to 16h starts in November and ends in February.

These experimental data indicate that the causative agency may not be the same for the northern and southern latitudes. One possible explanation is that corpuscular particles contribute to the production of ionization. The negative particles diffuse into the ionosphere in the northern latitudes, and the positive particles diffuse into the ionosphere in the southern latitudes. In the former, the electron density increases; in the latter, the positive ions combine with the electrons to produce neutral particles, thereby reducing the concentration of electrons. This oversimplified mechanism of a complex phenomenon suffers from insufficient data. A much larger sample of data must be analyzed in order to discover the true mechanism. Another suggestion is to look into the application of the drift theory, as treated so elegantly by Martyn.

*Daily variation at the equinox*—During the equinoctial months, March and September, the variation of  $f_oF2$  is alike. Basically, the same features are evident. The daily variation of the values of  $f_oF2$  is fairly flat for the most northerly stations and its amplitude increases with decrease in latitude. The maximum diurnal range of  $f_oF2$  in September 1956 and March 1957 is about 1 Mc/s for Thule and Resolute Bay, about 0.5 Mc/s for Kiruna and Tromsø, about 4.3 Mc/s for Baker Lake, and about 6 Mc/s for Fairbanks. It is 5.5 Mc/s in September and 7.5 Mc/s in March for Oslo. The nighttime values decrease with increasing latitude up to the

auroral belt, the more so in high sunspot years than in low sunspot years. They seem to be more or less equal within the auroral zone. Figure 2 reveals that every year a well-defined daily variation begins at Resolute Bay in September and ends in April. Farther south, at Oslo, for instance, it occurs one month earlier and lasts about one month longer.

The nighttime values of  $f_oF2$  lead to a morning dip, quite evident, but not so strongly marked as in middle latitudes. They rise to a maximum approximately at local noon. The data suggest a longitudinal effect for the time of occurrence of this daily maximum. The maximum occurs along the Scandinavian chain at 12h to 13h for Tromsø and Kiruna and at 14h for Oslo. On the Alaskan chain, it occurs, however, at 15h to 16h. At Tiksi Bay it occurs close to 17h. In Iceland and Canada the maximum occurs between those extrema. This difference cannot be overlooked, but we have no explanation for it.

*Daily variation in winter*—Winter  $F2$  characteristics contrast rather sharply with summer conditions. The most remarkable daily variation of  $f_oF2$  occurs during the winter months. The contrast between summer and winter is clearly shown in Figures 1 and 2. The former depicts the daily variation for the month of December, whereas the latter shows it to be a very regular phenomenon all through the winter. The amplitude changes quite considerably in the polar region; it is small or non-existent near the pole and increases rapidly southward. All over the polar regions the winter noon values of  $f_oF2$  are larger than the summer noon values. Such a behavior is found all over the globe and was first noted some twenty years ago by Appleton and Naismith.

At the highest latitudes, such as Longyearbyen, Figure 1 shows, for all practical purposes, no diurnal variation. While the median value is fairly constant over the 24 hour period, the individual values scatter from day to day over a fairly wide range of values (between 2.2 and 14.0 Mc/s in December 1956). This behavior could mean that no solar radiation reaches the  $F2$ -layer at  $78^\circ\text{N}$  in December to ionize the air. The ionization is carried into this region by atmospheric motions from an area south of this latitude.

This finding for Longyearbyen is corroborated by other data. Whatman [1949] made measure-

ments in Barentsburg/Svalbard in 1942-1943. No marked diurnal variation was found in December, when the diurnal variation amounted to about  $\pm 0.25$  Mc/s. *Driatsky* [1957] reported results from the drifting ice station SP-3. From May 1953 until April 1955 the station drifted from  $87^{\circ}\text{N}$  and  $180^{\circ}\text{E}$  to  $86^{\circ}\text{N}$  and  $325^{\circ}\text{E}$ . In winter, during December and January, when the sun did not illuminate the  $F_2$ -layer for about two months, ionization in the  $F_2$ -layer was still present; frequently the layer showed structure, spread  $F$  echoes, and clearcut triple magneto-ionic splitting. The critical frequencies,  $foF_2$ , varied over a wide range during this time; maximum values of 5 to 5.5 Mc/s occurred. He concluded that this variation was due to the existence of ionized clouds passing over the observation point.

In Thule, a slight daily variation occurs in December with a maximum in the afternoon hours. For Resolute Bay, a definite daily variation occurs in December, but the amplitude is small. But from Point Barrow southward, the amplitude is large, increasing with the distance of the station from the North Pole.

The nighttime value is followed by a dip around 07h and a steep rise. The maximum occurs around 13h, at about 12h to 13h along the Scandinavian chain and at 13h to 14h along the Alaskan chain, indicating some longitudinal effect as in the equinoctial months, but less pronounced. The rise and fall is rather symmetrical around the maximum, in marked contrast to the March-September curves and those for middle latitude stations. An evening dip occurs at some stations, but its reality is questionable. The two southern hemisphere stations exhibit the same daily variation during the southern winter. After attaining the maximum, the electron density decreases in an orderly fashion to about 20h, after which the fluctuations, characteristic of darkness, make their appearance. It is interesting to note, nevertheless, that in the winter of 1956 the nocturnal variations in critical frequency at Baker Lake were nearly 2.0 Mc/s, almost equal to the entire diurnal variation at the highest latitude station. Normally this variation is about 1 Mc/s or less.

It is of great interest to compute the hourly rate of increase of critical frequency after sunrise, as well as its hourly rate of decrease for December. It can reach (in extreme cases) about 3

Mc/s per hour. The equivalent net gain in electron density reaches large values; for example, 140 electrons/cm<sup>3</sup> sec between 08h and 09h in December 1956 for Oslo. As expected, close to the time of maximum, the rate of increase decreases sharply. The average hourly rate of increase, in critical frequency between the sunrise dip and noon, was computed by selecting the values between the interval 07h to 12h for all stations, except Baker Lake and Anchorage, where the interval 07h to 13h was chosen. For the average hourly decrease, the interval 13h to 18h was selected, except 14h to 20h for Resolute Bay and Oslo. The results are listed in Table 1 where the first column represents the rate of increase and the second column the rate of decrease.

TABLE 1—Average hourly rate of increase (I) of critical frequency in the morning hours and the corresponding rate of decrease (D) in the afternoon hours for December, data in Mc/s per hour

Station	December 1954		December 1955		December 1956	
	I	D	I	D	I	D
Thule	...	...	...	...	0.4	0.1
Resolute Bay	0.1	0.1	0.4	0.1	0.4	0.2
Tromsø	0.5	0.4	0.8	0.8	1.3	1.3
Fairbanks	(0.7)	0.7	0.9	1.4	(1.3)	1.3
Baker Lake	0.4	0.3	0.7	0.6	1.2	1.2
Anchorage	0.7	0.7	1.0	1.2	1.5	1.0
Oslo	0.7	0.5	1.2	1.0	2.2	1.6

These are *average* rates. For the maximum rates, on the other hand, differences between I and D exist and are obvious by a slight asymmetrical behavior of the curves in Figure 1. It is evident from the values in Table 1 that

- the rates are very small at Thule and Resolute Bay indicating a probable change in the ionizing agent for latitudes north of  $70^{\circ}$ , because only radiation scattered in the atmosphere or partially absorbed on the long path through the atmosphere can reach the  $F_2$ -layer from below at latitudes north of  $70^{\circ}$ ,
- for latitudes less than  $70^{\circ}$  geom. lat. the average rate of increase and the average rate of decrease are about equal, indicating a symmetrical behavior around noon

maximum; whereas, for Oslo, the rate of decrease is less than the rate of increase, and c. the rates I and D decrease with increasing latitude and for a given station increase with solar activity.

The maximum diurnal range in critical frequency is shown in Table 2. In terms of electron density, the diurnal variation in December may be close to a factor of 25, certainly an appreciable factor. The range increases southward and may exceed 10 Mc/s for the southern stations. This range increases strongly with the sunspot number, for in December 1956 this range is nearly three times as high as in December 1954.

TABLE 2—Maximum diurnal range for monthly median values of  $f_oF2$  in December in Mc/s. (If the minimum is based on insufficient data, the number appears in parentheses. The range is given to the nearest  $\frac{1}{2}$  Mc/s.)

Station	December 1954	December 1955	December 1956
Longyearbyen	...	...	(2.0)
Thule	...	...	(2.5)
Resolute Bay	1.0	2.0	3.5
Fromsøe	2.5	4.0	(8.0)
Kiruna	2.0	...	7.5
Fairbanks	(3.0)	(6.5)	(8.5)
Baker Lake	3.0	5.0	7.5
Anchorage	(4.5)	6.5	10.0
Oslo	4.0	6.5	11.0

*Other investigations*—Two important high-latitude series of data have been recorded in Svalbard, both of which fully support the conclusions we have drawn above for latitudes around 80°N. *Whatman* and *Hamilton* [1938] conducted ionospheric soundings at Brandy Bay in 1934–1935. Their results for local noon show a small annual variation of about 1 Mc/s, varying from 5.5 to 6.5 Mc/s, at a time when the sunspot number increased from 30 to 75. These  $f_oF2$  values agree very well with those for Longyearbyen in 1956. *Whatman* [1949] made is measurements in Barentsburg/Svalbard in 1942–1943 and published data for a four-month period. No marked diurnal variation was found in December and May, but some diurnal variation occurred in October and March. The general level of  $f_oF2$  was around 3–4.5 Mc/s for the monthly mean values in a sunspot minimum

period. The diurnal variation amounted to about  $\pm 0.25$  Mc/s in December and about  $\pm 0.50$  Mc/s in May. This agrees also with our findings for Thule and Longyearbyen. The maximum value of  $f_oF2$  in October 1942 reached only about 7 Mc/s; this is certainly a fairly low value for the highest possible frequency, but probably it is typical for a sunspot minimum condition in the inner Arctic.

*Annual variation*—While discussing the diurnal variation for the four seasons, we have already stated the main features and the peculiarities of the annual variation. Some additional information, however, seems warranted.

The pattern of the annual variation is clearly shown in Figure 2. It is basically similar at all polar stations and in agreement with the pattern established long ago for middle latitudes. The year 1954 reveals the general pattern best because the solar activity remained almost negligible throughout this year. Hence no change in the incoming solar radiation took place.

In this year, the area north of 70° geom. lat., as shown by Resolute Bay and Godhavn, obtained the highest values  $f_oF2$  in summer and not in winter. (This is also true for Longyearbyen in a high sunspot year.) South of 70° geomagnetic latitude the pattern is different, because the winter noon data are definitely the highest values within the year in agreement with middle latitudes. Large changes of  $f_oF2$  occur everywhere only around sunrise and sunset and can be compared with data on solar altitude at high latitudes (for example *Thomas* [1953]).

In a year of high sunspot number, for example 1956, the pattern has changed at the highest latitudes because, even at Thule, the maximal values occur always in winter around noon. The ratio of the winter noon to the summer noon attains values between one and two. It increases southward, and it also increases with solar activity, simply because the winter noon critical frequency follows this trend; whereas the summer noon values show very little variation with latitude, and they increase with solar activity only 20 to 50 pct. A closer inspection of the isopleths reveal that the isolines for the highest value  $f_oF2$  have a 'dumbbell' shape with a maximum core in November and another in February. For example, the maximal core values ( $f_oF2 > 13$  Mc/s) for 1956 occur around 60°N, but only 8 Mc/s appear over Thule and Resolute



Bay. The lowest values in the annual cycle occur during winter in the morning hours; they drop down to around 2 or 3 Mc/s depending on the sunspot activity. Inside the auroral zone this minimum value may be even higher, around 4 Mc/s during the sunspot maximum.

*Geographic and geomagnetic influence on  $F_2$ -layer*—In contrast to the obvious time-varying changes is the spatial variation in the ionosphere itself. The most important ones are variations of the critical frequency with latitude (geographic and/or geomagnetic) and changes in the vertical distribution of the electron density. Since a pronounced influence of the geomagnetic field on the critical frequency of the  $F_2$ -layer exists in the equatorial region, such an effect can be expected in the polar region too. Undoubtedly, some of the inadequate explanations used to elucidate the erratic behavior of the polar ionosphere have their origin in the paucity of very high latitude data.

The geomagnetic distortion of the  $F_2$ -layer has been extensively studied in low and middle latitudes; to name just one extensive study we refer to Maeda [1955]. Since no over-all investigation exists for the polar regions, we studied the relationship by plotting noon values of  $foF_2$  for many months as a function of geographic latitude, geomagnetic latitude, and dip angle, but no clear-cut relationship exists.

Finally, the noon data are plotted on a polar projection map (Fig. 3). Smoothed data are selected by computing the three-hour mean value  $(11 + 12 + 13)/3$ , called the 'noon' value. All arctic stations are used, and in addition, several stations located south of the polar circle, such as Inverness, Scotland; Adak, Alaska; and Ottawa, Canada. This southward extension is necessary in order to determine whether any singularity occurs at the auroral belt.

From our map collection we selected the months January to April 1954 to show the conditions at very low solar activity, and September to December 1956 to show the conditions at very high solar activity. The summer months are omitted because no large variation occurs. The difference between the noon data of all stations is so small, namely,  $4.25 \pm 0.25$  Mc/s in June 1954 and  $5.9 \pm 0.7$  Mc/s in June 1956, that it seems futile to look for a spatial effect at all. Certainly no difference between geomagnetic and geographic coordinates

can be found here. The fall to spring data are the best.

The actual stations used are shown as dots, but the station value has been left out, otherwise the figures would be too crowded. Lines of equal  $foF_2$  in steps of 0.5 or 1 Mc/s have been drawn through the observed data. In winter and spring 1954, the isolines follow fairly closely the geographic latitudes except over northwestern Europe where they are pushed towards the pole, indicating some slight geomagnetic effect.

The fall and winter maps for 1956 contain three important stations, Longyearbyen, Thule, and Resolute Bay, and the picture around  $80^\circ$  is certainly more reliable than that for 1954. On the other hand, some Canadian stations, Prince Rupert, B. C. and St. Johns, Newfoundland, have been closed. On the 1956 maps the auroral belt is added (after Vestine [1944]) as a thin dashed line. In this sequence,  $foF_2$  also decreases northward along a given longitude, but the isolines are 'pushed in' over Alaska and northwestern Europe, so that the lines of equal critical frequency fall between the geographic and geomagnetic latitude circles. This indicates a combined effect of solar radiation and the magnetic field, either by incoming charged particles from the sun or by an ionospheric drift under the influence of the earth's magnetic field.

In polar regions, the  $foF_2$  values can reach values of 8 Mc/s near  $80^\circ\text{N}$  and over 12 Mc/s at the auroral belt.

In November and December 1956 a steep gradient of  $d(foF_2)/d\phi$ , where  $\phi$  is the latitude occurs within the auroral zone. For December 1956 the values are 0.85 Mc/s/ $1^\circ$  latitude between Svalbard and Tromsø,  $\sim 0.6$  Mc/s/ $1^\circ$  latitude over North Alaska and  $\sim 0.27$  Mc/s/ $1^\circ$  over northern Canada. If  $\phi$  is in geomagnetic latitude then the values are 1, 1, and 0.3 Mc/s/ $1^\circ$  geomagnetic latitude respectively. This steep gradient around  $75^\circ\text{N}$  may actually be even more abrupt at the twilight zone in the  $F_2$ -layer. Results from the Ice Islands may help to clarify this point.

The results depict that at very high latitudes  $80^\circ\text{N}$ , ionospheric conditions differ markedly from those at  $70^\circ\text{N}$  or farther south. The maps clearly show the old difficulty of attributing the behavior of  $foF_2$  to one effect only, namely whether they are better explained when related to solar photon emission, (i.e., a function of the

geographic latitude) to magnetic effects (by charged particle current systems in or out of the atmosphere, by the static geomagnetic field, or a combination of both).

The only data recorded close to the North

Pole was published in a short note by Gassmann [1956]. In January 1955 he measured the critical frequency on a flight from Thule to the North Pole and back. During this flight,  $f_oF2$  stayed around 2 to 3 Mc/s, and the same level

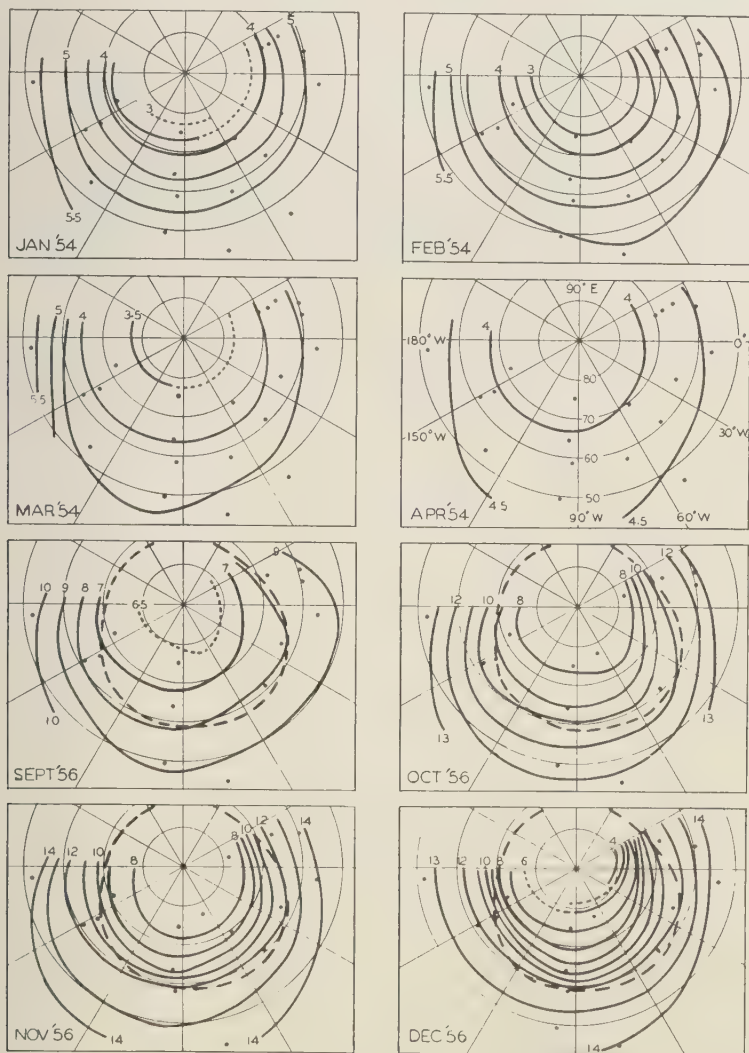


FIGURE 3

Polar maps of equal values  $f_oF2$  at noon for January - April 1954 and September - December 1956.

Example for years with low and high solar activity. The numbers mean  $f_oF2$  in mc/s. The dots depict the location of the stations used. The isochasm for maximal auroral activity is shown as a heavy dashed line for 1956 (after Vestine).



was measured simultaneously at Resolute Bay. When the plane arrived over the Pole at midnight (UT), the  $F$ -layer over the Pole and over Resolute Bay was in the earth's shadow. If this single example is typical, then the critical frequency between  $75^\circ\text{N}$  and the Pole remains fairly constant during the polar night.

As the IGY data from the arctic regions, including the Siberian and Russian coasts and the Ice Pack stations become available, this problem may be treated more successfully.

*Summary*—The critical frequencies for the  $F2$ -layer,  $foF2$ , are evaluated for polar stations. The data for the present solar cycle, namely 1954 to summer 1957, are taken from the CRPL publication. The temporal pattern is investigated and it is found that

1. The diurnal variation is largest in winter and smallest in summer.
2. In summer, the electron concentration changes very little during the day with an average critical frequency of between 5 to 7 Mc/s in high sunspot years and about 4 to 5 Mc/s in low sunspot years; the amplitude of the daily variation is of the order of 1 to 2 Mc/s or less.
3. In winter, the polar  $F2$ -layer displays a small diurnal variation at about  $75^\circ\text{N}$ ; there may be none farther north. The amplitude may reach 3 Mc/s at  $75^\circ\text{N}$  during a year of high solar activity. The amplitude increases southward up to 11 Mc/s at  $60^\circ\text{N}$  with a minimum around 7h and a maximum between 12h to 14h. The daily variation is fairly symmetrical around the maximum.
4. The pronounced daily variation lasts from fall through winter till late spring.
5. The annual variation for a given time of day shows a regular behavior with a maximum for the noon values in winter and a deep minimum for the morning and evening values also in winter.
6. In summer the latitudinal differences are small, of the order of 1 Mc/s or less between  $60$  and  $80^\circ\text{N}$ .
7. In winter, however, a very strong gradient exists, up to  $0.85 \text{ Mc/s/}1^\circ$  latitude. High absolute values occur at the southern border of the polar region, low values over the pole.
8. In years of high solar activity, the lines of equal critical frequency seem to fall between the geographic and geomagnetic latitude circles.
9. The diurnal curves for Deception ( $63^\circ\text{S}$ ) and Port Lockroy ( $65^\circ\text{S}$ ) show a dip around noon for the months of November to February.

*Acknowledgment*—The research in the document was supported by Air Force Cambridge Research Center under contract AF 19(604)-2630. The authors' appreciation is extended to Mr. Noel Stone, Air Force Cambridge Research Center, and to Mr. N. C. Gerson for many helpful discussions.

#### REFERENCES

- ALLEN, C. W., World-wide diurnal variations in the  $F2$  region, *J. Atm. Terr. Phys.*, **4**, 53-67, 1953.
- CHATTERJEE, B., Some regularities of the ionospheric  $F$ -region, *J. Geophys. Res.*, **58**, 353-362, 1953.
- DRIATSKY, V. M., Ionosphere in the region near the pole, paper presented at XI Gen. Assembly IUGG, Toronto, 1957.
- GASSMANN, G., Airborne ionospheric measurements in the north pole area, *J. Geophys. Res.*, **61**, 136-138, 1956.
- MAEDA, K. I., Geomagnetic distortion in the  $F2$  layer, *The physics of the ionosphere*, Phys. Soc., London, 245-253, 1955.
- MARTYN, D. F., A survey of present knowledge of the  $F2$  region, *The physics of the ionosphere*, Phys. Soc., London, 212-217, 1955a.
- MARTYN, D. F., Geomagnetic anomalies of the  $F2$  region and their interpretation, *The physics of the ionosphere*, Phys. Soc., London, 260-264, 1955b.
- RATCLIFFE, J. A., Some regularities in the  $F2$  region of the ionosphere, *J. Geophys. Res.*, **56**, 487-507, 1951.
- SKINNER, N. J., AND R. W. WRIGHT,  $F2$ -layer regularities at Ibadan, *J. Atm. Terr. Physics*, **5**, 290-297, 1954.
- THOMAS, M. K., Climatological Atlas of Canada, *Nat. Res. Council Can. no. 3151*, 188-189, 1953.
- VESTINE, E. H., The geographical incidence of aurora and magnetic disturbance, northern hemisphere, *Terr. Mag.*, **49**, 77-102, 1944.
- WHATMAN, A. B., Observations made on the ionosphere during operation in Spitsbergen in 1942-43, *Proc. Phys. Soc.*, London, **62B**, 307-320, 1949.
- WHATMAN, A. B., AND R. A. HAMILTON, High latitude radio observations, *Proc. Phys. Soc.*, London, **50**, 217-232, 1938.

(Manuscript received May 1, 1958.)

## Observations of Direction of Arrival of Long-Duration Meteor Echoes in Forward Scatter Propagation

T. HAGFORS AND B. LANDMARK

*Norwegian Defence Research Establishment,  
Kjeller, Lilliestrøm, Norway*

**Abstract**—The angular distribution of enduring meteor bursts in forward scattering is studied and compared with the angular distributions observed for the short duration meteor bursts and the turbulent background component. It is concluded that the enduring, fading bursts are due to meteor trails which reflect in a specular manner throughout the time during which they can be observed.

### INTRODUCTION

In recent papers by *Booker* [1956] and by *Booker and Cohen* [1956] a theory of long-duration meteor bursts was developed. According to this theory the initially straight columns of ionization left behind by meteoric particles entering the atmosphere are rendered rough by atmospheric turbulence soon after their formation. The theory is of great importance for the understanding of the complex phenomena involved in ionospheric forward scatter propagation. In ionospheric forward scattering there are two main modes of propagation, one is due to specular reflections from meteor trails and the other due to scattering from irregularities in electron density caused by atmospheric turbulence in the lower *E*-region.

In the papers referred to above it was shown that if meteoric ionization plays an important part in the signals observed, it is through the trails which have been rendered rough by the turbulence and not through the trails which are still capable of reflecting radio waves in a specular manner.

These points of view have been disputed by *Manning and Eshleman* [1957] who maintain that the long-duration fading meteor bursts can be adequately explained in terms of wind shears deforming the trails and thus permitting a reflection condition to be fulfilled at several points along the trail. As soon as multiple reflections occur, the fading of the echoes is readily explained.

It is the purpose of this paper to describe some observations of angle of arrival of long-duration meteor bursts as observed over a 46.8 Mc/s

scatter circuit between Tromsø (70°N, 19°E) and Kjeller (60°N, 11°E) and to show that the results may bear on the problem of how the trails are deformed.

### ANGULAR PROPERTIES OF THE SCATTER SIGNAL

In this section some angular properties of specular meteoric reflections and of the turbulent scatter signal as observed by us will be briefly reviewed in order to provide data for a comparison with the angular properties of the long-duration meteor bursts.

**Specular meteor reflections**—The specular reflections from meteor trails observed in forward scattering have been investigated theoretically [*Eshleman and Manning*, 1954; *Hines*, 1955] and experimentally [*Endresen* and others, 1958] by several workers. Due to the strict reflection condition, the angular distribution of the reflection points exhibits a pronounced minimum in the direction of the great circle path between receiver and transmitter and maxima at either side of this direction. On account of the rotation of the earth relative to the apex of the earth's path around the sun, the distributions are systematically assymetric in a manner depending on the time of day and on the orientation of the path. Two such distributions observed on the Tromsø-Kjeller circuit are shown in Figure 1. Details of the measurements are given by *Endresen* and others [1958].

**Turbulent scatter signal**—The angular distribution of the turbulent scatter signal is determined by the law of scattering, which in turn is related to the distribution of mean square

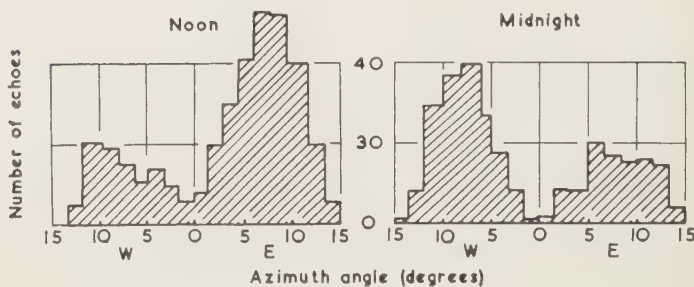


FIG. 1—Angular distributions of meteor trail reflections in forward scattering

electron density fluctuation among the various wave numbers in the spatial Fourier expansion of electron density. A method has been developed by one of the authors whereby it is possible to distinguish between the part of the Rayleigh-fading background signal which is due to specular reflections from meteor trails and the part which is due to turbulent scattering. The approximate angular power spectrum of the turbulent component observed on the Tromsø-Kjeller circuit is shown in Figure 2. It should be noted that this angular distribution is influenced partly by the (known) aerial polar diagrams and partly by the law of scattering.

#### ANGULAR DISTRIBUTION OF LONG-DURATION BURSTS

During November and December, 1957,

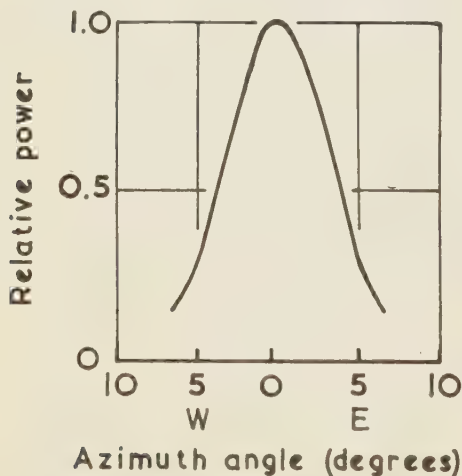


FIG. 2—Angular distribution of turbulence background signal

direction finder equipment was in operation at Kjeller in connection with a study of the angular distribution of specular reflections from meteor trails [Endresen and others, 1958] in forward scattering. The direction of arrival of the signal and the signal envelope in one of the aeriels were recorded simultaneously. The equipment covered an azimuth angle range of  $30^\circ$ , from  $15^\circ$  east to  $15^\circ$  west of the great circle path between Tromsø and Kjeller.

During the recording period a number of long-duration meteor bursts were observed and these have now been subjected to analysis. The bursts were divided into two classes, one consisting of bursts due to initially specularly reflecting trails and the other due to trails initially not appropriately orientated for reflection to occur. Examples of the two types of bursts are shown in Figure 3. The specular bursts are characterized by an abrupt beginning which is lacking in the other type of bursts.

The mean direction of arrival of individual bursts was read well after the fading had started. Since normally some fluctuation in the direction of arrival was observed, the angular fluctuation was also determined for individual bursts by inspection. In drawing the angular distribution this was accounted for by assigning a width to the bursts according to their angular fluctuation and plotting all bursts the same total area. The result of the analysis, which included 106 echoes, is shown in Figure 4.

#### DISCUSSION

It should be observed that the angular distribution of fading long-duration meteor echoes shown in Figure 4 is very similar in appearance to the angular distributions observed for short-



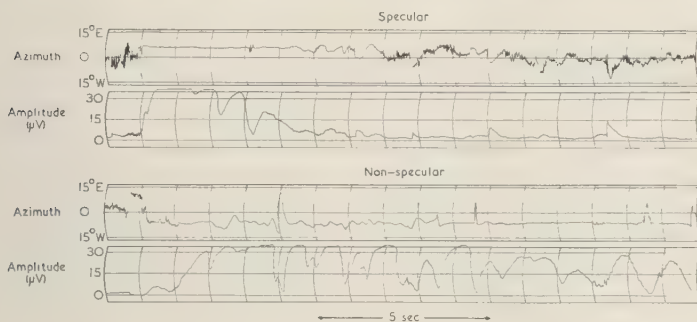


FIG. 3—Examples of two types of enduring meteor bursts

duration specular meteor reflections (see Fig. 1). Also it is interesting to note that echoes initially non-specular show a less pronounced minimum along the Kjeller-Tromsø path than that of the total distribution.

From these results it appears that the long-duration meteor bursts in forward scattering are highly aspect sensitive throughout the time during which they can be observed. It is also clear that the initially specularly reflecting trails are somewhat more aspect sensitive than the other type, but even these are aspect sensitive to a certain degree.

If the trails were broken up by atmospheric turbulence regardless of their orientation, they should be able to scatter the power towards the receiver. On this assumption one would expect the angular distribution of long-duration meteor

bursts to be similar to that of the turbulent scatter background signal. That this is not so is clearly evident from Figures 2 and 4.

The authors cannot, therefore, avoid the conclusion that the theory of long-duration meteor bursts based on atmospheric turbulence [Booker and Cohen, 1956] does not apply under the conditions of their experiment. A continuous background signal is, however, recorded during most of the time, particularly during the summer, and at least a part of this signal can be very well explained in terms of available theories on turbulent scattering. It is therefore suggested that turbulence might be more likely to develop in the height range usually taken to be responsible for the continuous background signal than in the height range in which most of the meteoric ionization takes place.

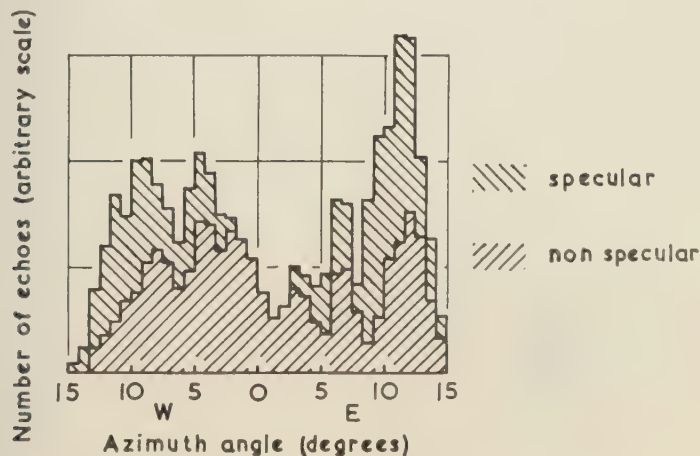


FIG. 4—Angular distribution of enduring meteor bursts

*Acknowledgment*—The work described in this paper is part of a research project sponsored by the Mutual Weapons Development Program.

## REFERENCES

- BOOKER, H. G., Turbulence in the ionosphere with applications to meteor-trails, radio-star scintillation, auroral echoes, and other phenomena, *J. Geophys. Research*, **61**, 673-705, 1956.
- BOOKER, H. G., AND R. COHEN, A theory of long-duration meteor-echoes based on atmospheric turbulence with experimental confirmation, *J. Geophys. Research*, **61**, 707-733, 1956.
- ENDRESON, K., T. HAGFORS, B. LANDMARK, AND J. RÖDSRUD, Observations of angle of arrival of meteor echoes in v.h.f. forward scatter propagation, *J. Atmospheric and Terrest. Phys.*, **12**, 329, 1958.
- ESHLEMAN, V. R., AND L. A. MANNING, Radio communication by scattering from meteoric ionization, *Proc. I.R.E.*, **42**, 530-536, 1954.
- HINES, C. O., Diurnal variations in the number of shower meteors detected by the forward-scattering of radio waves, *Can. J. Phys.*, **33**, 493, 1955.
- MANNING, L. A., AND V. R. ESHLEMAN, Discussion of the Booker and Cohen paper, "A theory of long-duration meteor-echoes based on atmospheric turbulence with experimental confirmation," *J. Geophys. Research*, **62**, 367-371, 1957.

(Manuscript received October 13, 1958.)

## Recurrent Geomagnetic Storms and Solar Prominences

RICHARD T. HANSEN

*High Altitude Observatory, Boulder, Colorado*

**Abstract**—A study is made of the run of areas of solar prominences around days of recurrent geomagnetic storms during the period 1917–1944. “Character figures for dark *H*-alpha flocculi,” taken from IAU Quarterly Bulletins of Solar Activity, are used as the measure of prominence areas. The earlier suggestions that the solar *M*-regions may be identified as quiescent prominences have not been confirmed.

**Introduction**—Several investigators have suggested that the hypothetical *M*-regions (the solar source of corpuscular radiation which may produce recurrent geomagnetic storms) are to be identified as the extended white light coronal streamers [Allen, 1944; Roberts and Pecker, 1955]. A limited number of observations of these white light streamers, at time of solar eclipse, shows them to be best associated with the quiescent type of prominences and to be directed away from active sunspot regions [Kiepenheuer, 1953; Van de Hulst, 1953]. The appropriate hypothesis has been advanced by Waldmeier [1946], by Kiepenheuer [1947], and by Bednarova [1953] that quiescent (solar) prominences should be associated with geomagnetic storms, and a number of statistical investigations have sought for this correlation.

Waldmeier [1946] compared the day-by-day areas of prominence profiles for 1930, and the character numbers of *H*-alpha filaments (from the IAU Quarterly Bulletins of Solar Activity), for April to December, 1943, with geomagnetic indices for the same periods. He concluded that the sequences of recurrent magnetic storms during these periods were related to the stationary prominences. Assuming the corpuscular flux to be directed radially from the sun in the coronal streamers, the time required for the particles to reach the earth was about 5.5 days.

Kiepenheuer [1947] measured the projected areas of filaments in a band  $14^\circ$  wide in heliographic longitude at the central meridian of the solar disk. He then used the superposed epoch method to study the variation of filament areas around magnetically disturbed and quiet days for a period of low solar activity, 1922–1924, and a period of high solar activity, 1926–1928. For 81 magnetically disturbed days in 1922–1924,

he found maxima of filament areas four days earlier, with the maxima about 10 pct above the mean values for the earlier period. The relation between filaments and magnetic activity was lost as solar activity increased.

Trotter and Roberts [1955] compared the average areas of prominences for the five most magnetically disturbed and the five most magnetically quiet days of each month in the periods May–November, 1951, February–October, 1953, and February–June, 1954. Using a superposed epoch analysis, they found no significant trend of prominence areas around either the magnetically disturbed or magnetically quiet days.

Leighton and Billings [1955] performed the same type of study as that by Trotter and Roberts for the same time period, but used measurements of disk-filament areas rather than limb-prominence areas. They concluded that the magnetic storms could not be associated with dark filaments in the period studied.

**Discussion**—The purpose of the present study is to examine uniformly the association of solar filaments with recurrent magnetic storms over a longer time period than has been attempted before. Also, we are interested specifically in the association of filaments with *recurrent* magnetic storms, rather than magnetic storms, per se. It should be noted that in some of the studies reported above, where the “most magnetically disturbed” days for each month were considered, the data sample was strongly diluted by sudden commencement, non-recurrent sequence magnetic storms. Statistical studies by Greaves and Newton [1928] and Allen [1944] give substantial evidence that two basic types of geomagnetic storms may be distinguished by their characteristics of recurrence, association with sunspots



and flares, storm intensity, and abruptness of commencement. We are here investigating recurrent magnetic storms. In the general case, such storms are non-sudden in commencement and are not associated with sunspots or flares.

The most well-defined recurrent geomagnetic storms from 1917–1944 were extracted from a list of *M*-region storms compiled by Trotter [1958]; many of these series were also noted earlier by Bartels [1932 and 1934]. A total of nine storm sequences, having between 4 and 20 rotations in each, are used in this study. The "Character Figures for Dark *H*-alpha Flocculi," published in the IAU Quarterly Bulletins of Solar Activity (from 1917–1944), is used as an index of solar filament areas. These character figures were assigned on a scale of 0–5, with the numbers referring to intensity and area of the dark hydrogen filaments for each day. The figure 0 was used to designate absence or extreme

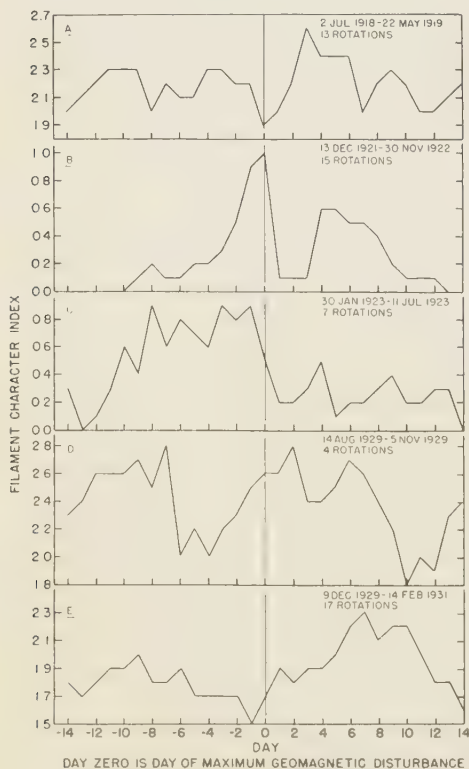


FIG. 1—The trend of solar prominence areas around days of recurrent geomagnetic storms (1918–1931).

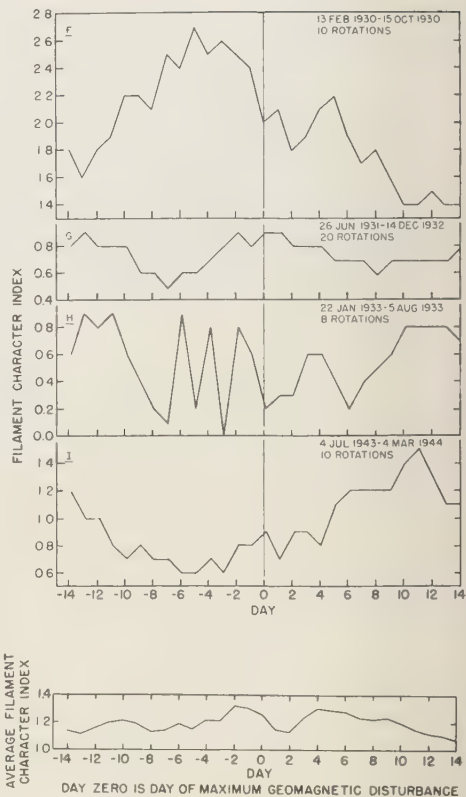


FIG. 2—The trend of solar prominence areas around days of recurrent geomagnetic storms (1930–1944) and, bottom, the average prominence areas for all nine storm sequences.

rarity; and 5, extreme abundance and intensity. Righini and Godoli [1950] correlated the character figures estimated from spectroheliograms taken at Arcetri and Mount Wilson with areas (uncorrected for foreshortening) of dark *H*-alpha flocculi measured on the same plates and expressed in units of the apparent disk of the sun. They concluded that character figures appear to be a function of areas alone.

In the first part of the study, a superposed epoch analysis was made of the run of dark filament indices around the approximate maximum day of recurrent geomagnetic disturbance for each of the nine storm sequences (Figs. 1 and 2). We see from the superposed epoch plots that in some cases, such as the sequence in 1923, there is a clear-cut maximum of filament areas

before the storm; in other cases, such as the sequence of 1944, there is a minimum of filament areas before the magnetic storm. The graphs show no consistent relation between filament areas and dates of recurrent magnetic storms, and the weighted average of all storm sequences is practically a flat curve.

Next we considered the run of geomagnetic index figures (Ap) around dates of high filament areas, using the same superposed epoch analysis technique (Fig. 3). For the years 1942-1944, zero days were chosen as the days of maxima in the trends of filament indices, provided the index values were at least 150 pct of the monthly mean. There were 59 such cases in the time period. No increase of geomagnetic activity is seen following days of large filament areas on the sun; the plots show, if anything, a slight decline.

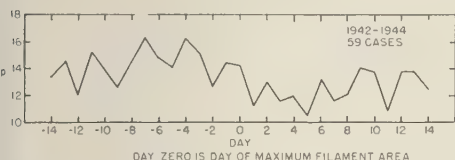


FIG. 3—The trend of daily geomagnetic indices around selected days of large areas of solar filaments.

**Acknowledgment**—This work was supported by the National Bureau of Standards (Contract NST-7019) as part of a general investigation of the sun's effects on geomagnetism. I wish to thank Miss Sylvia Moscové, Dr. Constance Warwick, and Mr. Robert Brun for their large contributions to the study.

# REFERENCES

- ALLEN, C. W., Relation between magnetic storms and solar activity, *Monthly Notices Roy. Astron. Soc.*, **104**, 13-21, 1944.
- BARTELS, J., Terrestrial magnetic activity and its relation to solar phenomena, *Terrestrial Magnetism and Atmospheric Elec.*, **37**, 1-52, 1932.
- BARTELS, J., Twenty-seven day recurrences in terrestrial-magnetic and solar activity, 1923-1933, *Terrestrial Magnetism and Atmospheric Elec.*, **39**, 201-202, 1934.
- BEDNAROVA, B., *Travaux de l'institut geophysique de l'Academie tchecoslovaque des sciences*, No. 11, 1953.
- GREAVES, W. M. H., AND H. W. NEWTON, Large magnetic storms and large sunspots, *Monthly Notices Roy. Astron. Soc.*, **88**, 556-567, 1928.
- GREAVES, W. M. H., AND H. W. NEWTON, Recurrence of magnetic storms, *Monthly Notices Roy. Astron. Soc.*, **89**, 641-646, 1929.
- KIEPENHEUER, K. O., A slow corpuscular radiation from the sun, *Astrophys. J.*, **105**, 408-423, 1947.
- KIEPENHEUER, K. O., *The Sun*, Ed. G. P. Kuiper, University of Chicago Press, p. 417, 1953.
- LEIGHTON, H., AND D. BILLINGS, *J. Atmospheric and Terrest. Phys.*, **7**, 349, 1955.
- RIGHINI, G., AND G. GODOLI, The physical meaning of the character figures of solar phenomena, *J. Geophys. Research*, **55**, 415-422, 1950.
- ROBERTS, W. O., AND J. C. PECKER, Solar corpuscles responsible for geomagnetic disturbances, *J. Geophys. Research*, **60**, 33-44, 1955.
- TROTTER, D., AND W. ROBERTS, *J. Atmospheric and Terrest. Phys.*, **6**, 282, 1955.
- TROTTER, D., High Altitude Observatory Solar Research Memorandum 107, June 1958.
- VAN DE HULST, *The Sun*, Ed. G. P. Kuiper, University of Chicago Press, p. 295, 1953.
- WALDMEIER, M., An attempt at an identification of the M-regions, *Terrestrial Magnetism and Atmospheric Elec.*, **51**, 537-542, 1946.

(Manuscript received September 9, 1958)





## An Analysis of Drifts of the Signal Pattern Associated with Ionospheric Reflections

DONALD G. YERG

*Michigan College of Mining and Technology,  
Houghton, Michigan*

**Abstract**—If the signal pattern in the vicinity of three spaced receivers is assumed to be represented by a linear variation in distance and time for a short time interval  $\tau$ , the apparent motion of the pattern must be normal to the lines of constant signal intensity. The instantaneous velocity of the ground pattern may be represented as the sum of the component of the average drift velocity in the direction of the gradient and a random velocity in the direction of the gradient. After the drift velocity has been determined by correlation analysis, the random velocity for each time interval may be obtained, and a spectral analysis of the random speed may be performed. This information may be used as an aid in describing the detailed motion of the signal pattern which is necessary in relating the observations to ionospheric motions.

**Introduction**—Radio reflections from the ionosphere may be obtained on receivers located at the corners of a triangle. The three signal intensity-time records may be analyzed to yield the average velocity of the ground pattern of signal intensity. The ionospheric drift velocity is assumed to be one half the velocity of this pattern. In the analysis, the fluctuations in the reflected signal are attributed to the drift of the signal pattern, unchanging in mean size and shape during the interval required to cross the receiver site, and to random fluctuations superimposed on this pattern. The reduction of the data yields both a drift velocity and a characteristic velocity which is a measure of the fading due to random changes.

In the effort to use this type of observation for studying ionospheric motions, it is desirable to obtain as much information as possible concerning the steady and random motions of the signal pattern. In this paper, a statistical procedure is developed for studying the detailed motion of the pattern. Particular attention is given to the description of the random motion and the orientations of the pattern.

**Drift Velocities and Random Velocities**—The three receivers are placed at the corners of a right triangle. The receiver separation may be assumed sufficiently small, so that the instantaneous signal may be represented by a linear variation during the interval in which any given isopleth of signal intensity is in the immediate vicinity of the receivers. Thus,

$$R(x, y, t + \tau) = R(0, 0, t) + (\partial R / \partial x)x + (\partial R / \partial y)y + (\partial R / \partial t)\tau \quad (1)$$

where  $R$  is the observed signal intensity. The total derivative of  $R(x, y, t)$  with respect to time may be put in the form

$$dR/dt = \partial R / \partial t + \vec{V} \cdot \nabla R \quad (2)$$

where  $V$  is the constant vector representing the drift velocity and  $dR/dt$  represents the change in structure resulting only from random changes. With respect to a moving coordinate system, this equation becomes

$$dR/dt = \delta R / \delta t + \vec{V}_i \cdot \nabla R \quad (3)$$

where  $V_i$  is the drift velocity relative to the moving coordinate system, and  $\delta R / \delta t$  is the time variation observed in the moving coordinate system. Subtraction of (3) from (2) shows that

$$\delta R / \delta t = \partial R / \partial t + (\vec{V} - \vec{V}_i) \cdot \nabla R \quad (4)$$

Let

$$\vec{V}_L = \vec{V} - \vec{V}_i \quad (5)$$

Then,  $V_L$  is the velocity of the moving coordinate system. If  $V_L$  is selected in any direction with a speed such that  $\delta R / \delta t = 0$ , the speed is found from (4) and (5) to be

$$V_L = -(\partial R / \partial t)(\partial R / \partial L)^{-1} \quad (6)$$

where  $\partial R / \partial L$  is the derivative in the direction of  $V_L$ . Thus,  $V_L$  is the apparent velocity of a



experiment [Yerg, 1956]. That these averages are zero is the natural consequence of the assumption that changes in structure of the drifting pattern are random changes, and this conclusion is demonstrated by averaging the components of the random velocity.

The components of  $V'$  along the coordinate axes may be found by multiplication with the sine and cosine of the angle between the normal and the  $X$ -axis. Thus,

$$V_x' = (dR/dt)(\partial R/\partial x)(\partial R/\partial n)^{-2} \quad (15)$$

$$V_y' = (dR/dt)(\partial R/\partial y)(\partial R/\partial n)^{-2} \quad (16)$$

The average values of these components must equal zero. Since the denominators are always positive and are independent of the numerators, it follows that

$$\overline{(dR/dt)(\partial R/\partial x)} = \overline{(dR/dt)(\partial R/\partial y)} = 0 \quad (17)$$

This is the condition necessary for the evaluation of  $V_x$  and  $V_y$  from a stationary coordinate system.

After  $V_x$  and  $V_y$  have been determined from equations (13) and (14) or from the equivalent correlation analysis,  $dR/dt$  can be determined from (2) for each time,  $t$ , in a record of given length, assuming linear variations of  $R(x, y, t + \tau)$ . Then, the corresponding values of  $V_x'$  and  $V_y'$  can be found from (15) and (16). In this manner,  $V'$  is found as a function of time.

**Computation of Derivatives**—It is assumed that the derivatives required in the preceding equations may be approximated by finite differences in the records. However, such computations require absolute measurements of the signal intensity, and these are not available. Let it be assumed that the absolute value of the signal intensity,  $R$ , is a linear function of the observed signal,  $r$ . Then,

$$R(0, 0, t) = a_{0,0}r(0, 0, t) + b_{0,0} \quad (18)$$

$$R(x, 0, t) = a_{x,0}r(x, 0, t) + b_{x,0} \quad (19)$$

Also, the statistical description of the absolute signal intensity should be independent of the choice of origin, so that the mean and variance at the two locations may be assumed equal. This yields

$$\begin{aligned} a_{0,0}\overline{r(0, 0, t)} + b_{0,0} \\ = a_{x,0}\overline{r(x, 0, t)} + b_{x,0} \end{aligned} \quad (20)$$

$$\begin{aligned} a_{0,0}^2 \overline{[r(0, 0, t) - \overline{r(0, 0, t)}]^2} \\ = a_{x,0}^2 \overline{[r(x, 0, t) - \overline{r(x, 0, t)}]^2} \end{aligned} \quad (21)$$

$$a_{x,0}/a_{0,0} = \sigma_{0,0}/\sigma_{x,0} \quad (22)$$

where  $\sigma_{0,0}$  and  $\sigma_{x,0}$  are the standard deviations of the observed signals at the two locations. Then, with the aid of equations (18) to (22), it follows that

$$\begin{aligned} \partial R/\partial x &= (a_{0,0}\sigma_{0,0}/\Delta x) \\ &\cdot \{[r(x, 0, t) - \overline{r(x, 0, t)}]\sigma_{x,0}^{-1} \\ &- [r(0, 0, t) - \overline{r(0, 0, t)}]\sigma_{0,0}^{-1}\} \end{aligned} \quad (23)$$

where  $\Delta x$  is the distance between the two receivers. Thus, the derivative can be determined from the observed data except for the unknown constant.

In a similar manner, the other derivatives can be determined except for the common factor,  $a_{0,0}$ . In the evaluation of the random velocities, ratios of derivatives are required, so that the constant,  $a_{0,0}$ , divides out of the equation and does not appear in the computation. The random velocities are thus determined from the observed signals and their respective means and standard deviations.

**Analyses of Data**—Eight sets of data were obtained through the courtesy of T. N. Gautier of the Upper Atmospheric Research Division, CRPL, National Bureau of Standards. These observations were made in 1950 in Washington, D. C. [Salzberg and Greenstone, 1951].

The records were scaled at intervals of one half second and the drift velocities were computed by the graphical correlation method [Briggs and others, 1950] at CRPL in Boulder, Colorado. The data were then made available to the author for further analysis.

The first two sets of data were analyzed with the aid of desk calculators, and subsequently the remaining six were evaluated with an IBM 650 digital computer at the IBM Laboratories in Poughkeepsie, New York, by K. McMillin of the Mathematics Department of the Michigan College of Mining and Technology. These data illustrate the types of information obtainable by the above statistical procedure.

Figure 2 shows the histograms of random speeds. The speeds were not divided by two, and so the values refer to the ground pattern



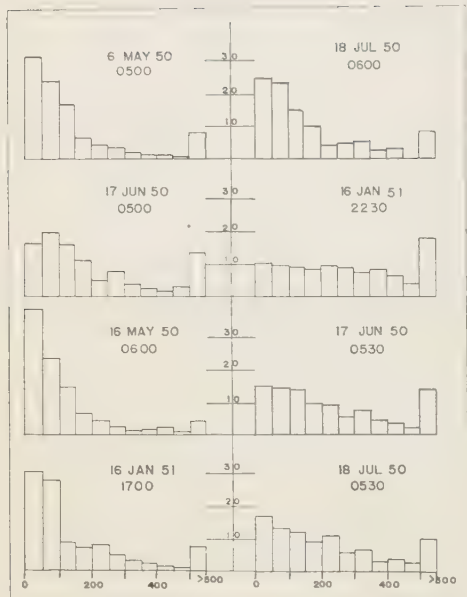


FIG. 2—Histograms of random speeds (frequencies are in per cent and speeds are in m/sec)

rather than ionospheric irregularities. Considerable variation in the distribution of speeds occurs, although 85 per cent of the speeds are less than 500 m/sec in all but one example.

Figure 3 shows the frequency diagrams for directions of the random velocity. Since these directions were taken normal to the isopleths of signal intensity, the diagrams also describe the orientation of the signal pattern. The asymmetry of the diagrams is striking. Roughly, the most frequent directions are approximately 180 degrees apart. This suggests the likelihood of a nearly ruled pattern undergoing random changes of the spacing between contours as the pattern drifts across the receiver site.

If a ruled pattern exists, equations (13) and (14) are not independent equations. Only the normal velocity can be obtained. This also is true of the correlation methods. The validity of the drift computations depends upon the presence of sufficiently many different orientations to assure proper averages of the products of the derivatives. The random velocities computed here are to be regarded as departures from the average velocity vector. If the presence of a nearly ruled pattern leads to erroneous

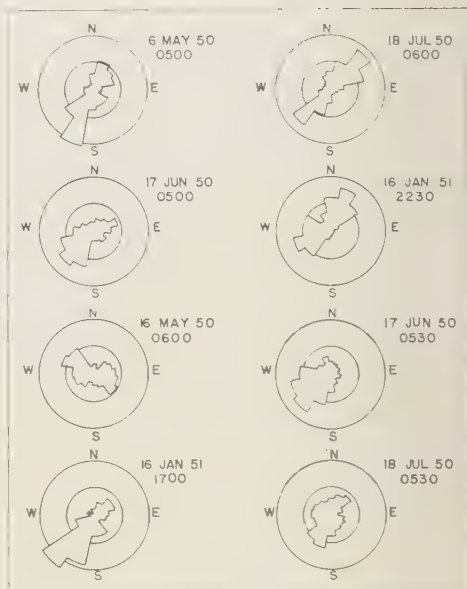


FIG. 3—Frequency diagrams of directions of random velocity reported for a 16-point compass (frequency contours are at 10 per cent intervals; velocity vectors are directed towards direction shown)

values of the drift velocity, the physical significance of the random velocities is also in question. Procedures are now being developed to check these interrelations further.

The random speeds were averaged for each direction of a 16-point compass and were plotted on polar diagrams. No two of the eight patterns were comparable in essential details, although there was some tendency for average speeds in opposite directions to be similar in value. Such diagrams are useful in describing the motion of the pattern, but a greater number of observations are needed.

Figure 4 shows the random velocity plotted as a function of time for the record of May 6, 1950 at 0500. Most frequently the vector is in the NE or SW quadrant as would be expected from the corresponding diagram in Figure 3. The curves show that the velocity vector oscillates between these two quadrants and again suggest a ruled pattern undergoing fluctuations of the contour spacing. Rather large velocities appear at more or less periodic intervals and are superimposed on smaller oscillations. The larger velocities are an inherent result of the method of

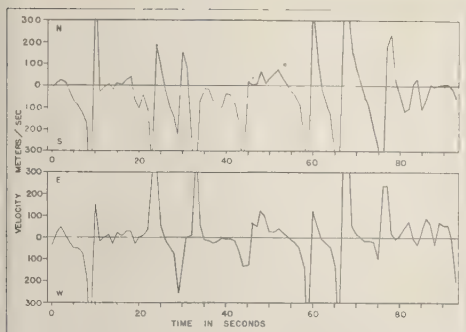


FIG. 4—The components of the random velocity as a function of time

approximating the derivatives. Such velocities occur when the gradient is near zero and the computation equations are not strictly applicable.

Figure 5 shows spectral curves for the record of May 6, 1950 at 0500 and June 17, 1950 at 0500. The maximum contribution to the sum of the squares of the speeds is associated with oscillations with periods in the neighborhood of 5.5 seconds. A small secondary maximum is associated with periods in the vicinity of 1.4 seconds and probably corresponds to the large fluctuation in velocities shown in Figure 4.

*Summary*—The preceding types of data analysis provide considerable information about the detailed motion of the signal pattern. The asymmetry of the frequency diagram of directions is more revealing of the structure of the pattern than is the orientation and eccentricity of the ellipse resulting from correlation analysis. The

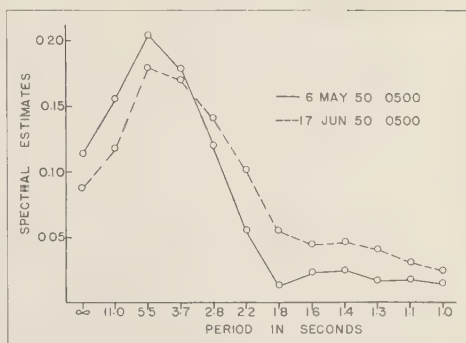


FIG. 5—Spectral estimates of random velocity—time curve

distribution of random speeds and the corresponding spectral analysis also gives a more complete insight concerning the significance of random processes. It is expected that more extensive analysis of these types will provide background information useful in the study of physical processes in the ionosphere associated with radio fading.

#### REFERENCES

- BRIGGS, B. H., G. J. PHILLIPS, AND P. H. SHINN, *Proc. Roy. Soc.*, **B63**, 106, 1950.  
 SALZBERG, C. D., AND R. GREENSTONE, Systematic ionospheric winds, *J. Geophys. Res.*, **56**, 521-533, 1951.  
 YERG, D. G., Notes on correlation methods for evaluating ionospheric winds from radio fading records, *J. Geophys. Res.*, **60**, 173-185, 1955.  
 YERG, D. G., *J. Atm. and Terrest. Phys.*, **8**, 247, 1956.

(Manuscript received September 13, 1958)





## Summer Upwelling Along the East Coast of Florida

C. B. TAYLOR AND H. B. STEWART, JR.

*U. S. Coast and Geodetic Survey, Washington, D. C.*

**Abstract**—Daily sea-water temperature measurements made at tide stations by the U. S. Coast and Geodetic Survey reveal anomalously low water temperatures during most summers at stations along the northern half of the east coast of Florida. The periods of colder water are accompanied by a lowering of sea level and by winds from directions most conducive to producing upwelling. Examination of daily variations in these factors at Daytona Beach during a year when the summer drop in mean water temperature exceeded  $10^{\circ}\text{F}$  shows the dependence of upwelling on the prevailing winds and suggests that the upwelling mechanism has no momentum but ceases at the coast as soon as the driving force is terminated.

**Introduction**—Sea-water temperatures measured along the east coast of Florida at Daytona Beach and Canova Beach (Fig. 1) reveal an apparently anomalous decrease in temperature during the summer months (Fig. 2). North and south of these stations the curve of mean monthly sea-water temperatures rises smoothly from a low in January to a maximum during July and August and then decreases during the fall to the January low. At Daytona Beach and Canova Beach mean sea-water temperatures rise from the January low to a high in June; however, lower values are found at these stations during July and August. The curve then rises to a second high in September and decreases again to the January low. The curves in Figure 2 are based upon the full series of observations available at each station. The occurrence of this anomaly at Daytona Beach has been pointed out by Green [1944] and Wagner [1957a].

The purpose of this paper is to examine the areal extent and magnitude of this summer decrease in coastal water temperature, to compare it with other measured variables, and to attempt an explanation of the anomaly in terms of coastal upwelling resulting from summer changes in the wind pattern.

Water-temperature measurements obtained at U. S. Coast and Geodetic Survey tide stations were made to the nearest half degree in samples drawn by bucket from a foot or two beneath the surface. Density was measured by hydrometer, and values have been reduced to the density at a standard temperature of  $15^{\circ}\text{C}$  ( $59^{\circ}\text{F}$ ). The observations usually were made once each week day, and it may be assumed that in the course of a month or a year the distribution of observations

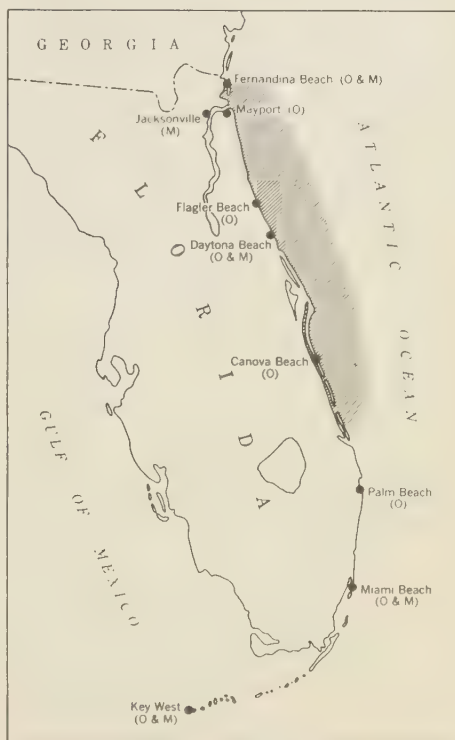


FIG. 1—Location of stations where oceanographic (O) and meteorological (M) observations were made; shading shows approximate area of summer temperature anomaly; its seaward extent is unknown

is fairly uniform over all phases of the tide, thus eliminating tidal influence from the mean values. At Daytona Beach and Canova Beach, water temperatures during some years were also

obtained by continuously recording thermographs. The temperature and density data on which this study is based have been presented elsewhere [*U. S. Coast and Geodetic Survey*, 1955, 1957]. Air temperatures and wind data were supplied by the U. S. Weather Bureau. The location of stations at which oceanographic and meteorological observations were made is shown in Figure 1.

*Magnitude and areal extent*—From the curves

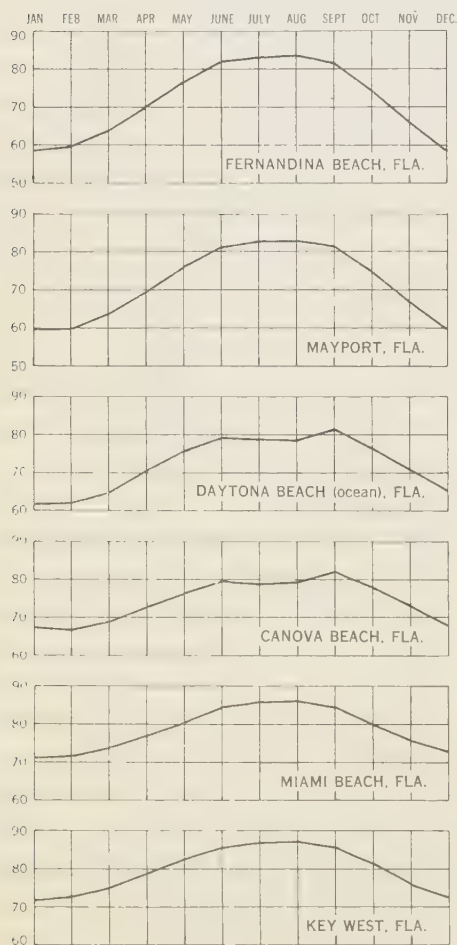


FIG. 2—Monthly mean surface-water temperatures (°F) Fernandina Beach to Key West, Florida, showing summer temperature decrease at Daytona Beach and Canova Beach; based on available data 1925–1954

in Figure 2, it is seen that at Daytona Beach and Canova Beach the summer dip in mean temperature is as much as 5° below an inferred undisturbed curve. Because these curves represent means derived from different numbers of years for each station, and because the phenomenon does not occur during every year nor during the same month each year, a more exact estimate of the magnitude of this temperature anomaly may be gained by examining the values for specific years. Table 1, therefore, lists the

TABLE 1—Monthly mean surface sea-water temperatures for summer months, °F

Year	Temperature at							
	Daytona Beach				Canova Beach			
	June	July	Aug.	Sep.	June	July	Aug.	Sep.
1940	76.1	80.4	83.3	81.0				
1941	76.6	75.9	74.8	82.0				
1942	79.5	79.0	75.9	81.9	Station established in 1946			
1943	83.3	80.8	80.2	82.8				
1944	84.2	83.3	80.5	83.2				
1945	77.5	79.2	80.4	81.7				
1946	79.5	82.6	80.8	82.6	77.7	79.2	76.8	82.4
1947	79.3	79.5	81.5	83.1	78.1	77.0	80.1	81.5
1948	81.7	81.3	79.4	82.2	78.9	77.2	77.3	82.1
1949	78.6	81.3	77.0	80.8	80.6	81.4	76.5	81.0
1950	80.4	70.3	78.6	77.8	79.7	72.6	78.1	78.7
1951					81.0	79.1	81.7	83.7
1952					79.4	81.3	79.0	83.3
1953	Station destroyed by hurricane in October 1950				80.0	80.9	81.4	82.3
1954					80.7	80.1	82.8	83.1
1955					78.3	81.9	84.7	83.9
1956					81.0	72.3	80.4	...
1957					78.4	77.4	80.5	82.1
Avg.	79.7	79.4	79.3	81.7	79.5	78.4	79.9	82.2

monthly means for summer water temperatures at Daytona Beach and Canova Beach from observations available from 1940 to 1957. The largest temperature decrease was observed in 1950 when July averaged 10.1°F cooler than June at Daytona Beach. During only four years of the 18-year period was a decrease in water temperature not found during July or August at one of these stations.

The distribution of colder summer water as observed at other stations along the coast is shown graphically in Figure 3. From this it appears that the greatest change occurs in the area of Canova Beach and Daytona Beach,

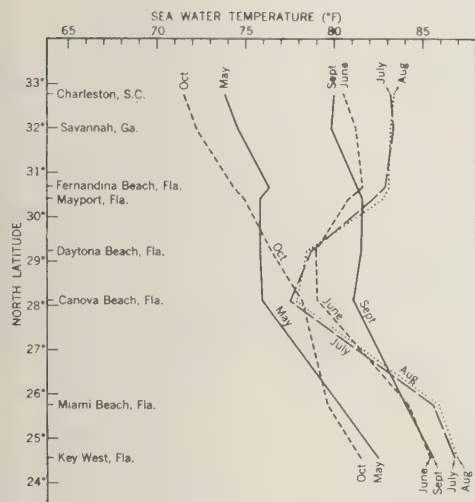


FIG. 3—Monthly mean surface-water temperatures plotted to show geographic extent of summer decrease along the Florida east coast; based on available data 1925-1951

decreasing both to the north and to the south. The one year of available data at Flagler Beach, 19 mi north of Daytona Beach, shows a July decrease only slightly less than at Daytona Beach. The anomaly is not apparent in the mean values at Mayport, Florida, 90 mi north of Daytona Beach nor at Miami Beach, 170 mi south of Canova Beach (Fig. 2). This, however, possibly presents a somewhat erroneous picture of the actual extent of the condition, for the stations at Mayport and Fernandina Beach are not on the open coast but located slightly inland on tidal rivers in a position not as readily accessible to changes in the ocean water temperature as are the other stations to the south. That fluctuations in the temperature of the ocean water are less apt to occur at these somewhat protected stations is at least suggested by the data for 1950. This year showed the greatest recorded decrease at Daytona Beach and was one of only two years from 1944 to 1957 in which summer decreases were also recorded at Mayport and Fernandina Beach. It is possible also that this was the only year in which the phenomenon spread that far north. In either case, the summer temperature decrease has not been observed at any station north of Fernandina Beach, indicating this as the northern limit.

To the south, Palm Beach had a station in operation only during 1946, and no summer decrease was noted there, while Daytona Beach and Canova Beach showed August means of  $1.8^{\circ}\text{F}$  and  $2.4^{\circ}\text{F}$  below their July values. Farther south, Miami Beach from 1940 to 1951 showed a summer temperature decrease in coastal waters only in July 1943 and in the limited observations of July 1950. Key West at the southern tip of the state shows summer decreases in only three of the 13 years for which records are available since 1940.

The areal extent of the summer cooling of coastal waters may also be delineated by showing the mean water temperature difference between adjacent stations by months (Table 2). Note

TABLE 2—Differences ( $^{\circ}\text{F}$ ) in mean monthly sea-water temperature between adjacent stations, Southport, N. C., to Key West, Florida, based on 8 to 32 years of observations; a negative value means the more southerly station had colder water

Temperature difference between	March	April	May	June	July	Aug.	Sep.
Charleston and Southport	2.2	1.5	0.7	1.3	1.2	1.3	2.1
Savannah and Charleston	1.4	1.1	0.9	0.9	0.3	0.3	0.0
Fernandina and Savannah	4.9	3.2	1.5	0.2	-0.5	-0.5	1.1
Mayport and Fernandina	0.0	-0.3	-0.4	-0.7	-0.7	-0.6	0.3
Daytona and Mayport	1.0	1.6	0.5	-1.4	-3.1	-3.3	0.3
Canova and Daytona	3.9	1.3	-0.5	-0.3	-0.8	-0.4	0.2
Miami and Canova	4.9	4.9	4.3	4.9	7.1	7.2	2.4
Key West and Miami	1.2	1.6	2.2	1.2	1.1	0.9	1.2

that the largest summer temperature change occurs between Daytona Beach and Mayport on the north and between Canova Beach and Miami Beach on the south.

Thus the colder summer temperatures are most pronounced in the Canova Beach-Daytona Beach area and are not found north of Fernandina Beach. To the south, the phenomenon probably dies out between Canova Beach and Palm Beach except during periods of extraordinarily large decreases in water temperature when the effects have been noted as far south as Key West.



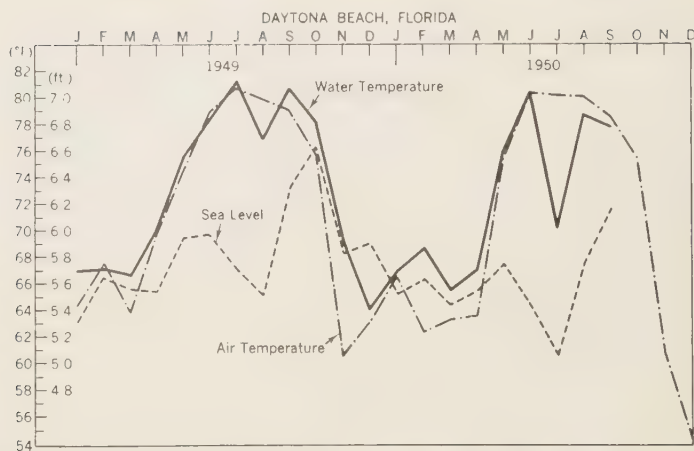


FIG. 4—Relationship between sea water temperature, sea level (feet above staff zero), and air temperature at Daytona Beach, 1949–1950

The seaward extent is unknown, as the data are from coastal stations only.

*Sea-level variations*—If this temperature decrease is caused by the upwelling or exposure of colder water after surface water has been blown seaward, then it might be expected that a lowering of sea level would accompany the occurrence of the colder water. When sea-level values and temperature values were compared, it was found that the annual curves for these two factors were very similar in shape in the area under consideration. Figure 4 shows the annual variation in sea level and water temperature at Daytona Beach for the two-year period 1949–1950. A marked summer decrease in sea level is noted in both years coinciding exactly with the summer decrease in water temperature. This summer lowering of sea level is not limited to the Daytona Beach area, since over a long period of observations the stations along the Atlantic Coast south of Chesapeake Bay all show a secondary minimum in July [Marmer, 1951]. This July minimum is most pronounced, however, along the northern part of the east coast of Florida.

This lowered sea level could in part be due to steric changes resulting from the lower water temperature [Pattullo and others, 1955]. Only surface-water data are available however, so no attempt could be made to compute the steric variations. However, the major portion of the decrease in sea level is believed to be the direct

reflection of the summer change in wind pattern discussed below. The effect of winds on sea level along the east coast has been well documented for South Carolina by DeVeaux [1955] and for Atlantic City and Nantucket Sound by Miller [1954, 1958].

*Density variations*—The density of the surface sea-water samples depends both on the salinity and the temperature, increasing with increasing salinity and with decreasing temperature. Observed density values were first reduced to a standard temperature of 15°C, in order to eliminate the effect of the water temperature, and were then converted to the salinity values corresponding to these densities. When mean salinities for the Florida east coast [U. S. Coast and Geodetic Survey, 1957] were examined, Mayport and Fernandina Beach showed uniformly lower salinity, reflecting the location of these stations near fresh water outflow as mentioned previously. At the other stations where lower summer temperatures were found, the average salinity range was only 0.5 ‰ from April to September. Though higher salinities might be expected with upwelling, the range of observed salinities is so small that the evidence is inconclusive.

*Surface winds*—With little information available on annual changes in the coastal currents off the Florida coast, the surface winds were examined directly as a possible cause for the observed summer temperature anomaly. It is

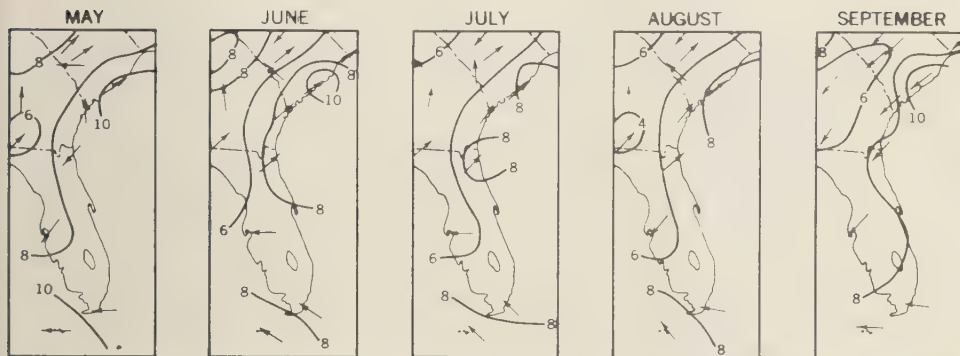


FIG. 5—Average hourly velocity and prevailing direction of wind, May through September [U. S. Weather Bureau, 1942]

well known that wind-generated currents tend to flow to the right of the wind in the northern hemisphere [Ekman, 1905]. It has also been demonstrated both theoretically and by observation that coastal upwelling of colder, more dense water may result when winds of sufficient strength and duration blow from directions that tend to move the surface waters away from the coast [Sverdrup, 1938; Sverdrup and Fleming, 1941; Defant, 1952; Longard and Banks, 1952; Hidaka, 1954].

Charts showing the average hourly surface wind velocity and the prevailing direction have been compiled by months from long series of observations by the U. S. Weather Bureau [1942]. Portions of these charts covering the Florida coast are shown for May through September in Figure 5. Prevailing winds are defined as the most frequent wind direction during the day, and average velocity as the average of the hourly velocities regardless of direction. Thus some directions are ignored and no weight is given to the hours with the highest velocities. Taken over a period of years, however, the mean values should be representative. From Figure 5 it is evident that along the northern half of the east coast of Florida winds blow predominantly onshore during May and September and offshore during June, July, and August, with highest average velocities in these latter months occurring in the general region of Daytona Beach. In the Miami Beach area, on the other hand, the prevailing direction is onshore throughout the period.

Since Daytona Beach is in the general area

having highest average offshore wind velocities and had the greatest summer decrease in water temperature in any one year (July mean  $10.1^{\circ}\text{F}$  colder than June in 1950), the effect of the wind was examined in greater detail at the Daytona Beach station for that year. The coast at Daytona Beach trends NNW-SSE. Daily values for the prevailing wind direction and average velocity were resolved into two components: one parallel to the coast and one perpendicular to the coast. The component parallel to the coast was considered positive toward the direction (NNW) that would tend to move water offshore, and the component perpendicular to the coast was considered positive also toward the direction (ENE) more likely to move water away from the coast. The two components for each day were added algebraically, and the sum therefore indicates by its sign whether the net effect for that day was to move the surface water toward (negative) or away from (positive) the coast.

Values for the wind effect at Daytona Beach were plotted for each day from May 1 to September 1, 1950, together with daily values for the surface water temperature, sea level, and air temperature (Fig. 6). From May 1 to June 14 the winds (Fig. 6a) vacillated back and forth in their tendency to move water toward or away from the coast. Then for the more than seven weeks from June 14 until August 6 the wind effect was dominantly positive, that is, tended to move water away from the coast, and a negative (onshore) wind effect was noted on only eight days during this period. On August 6

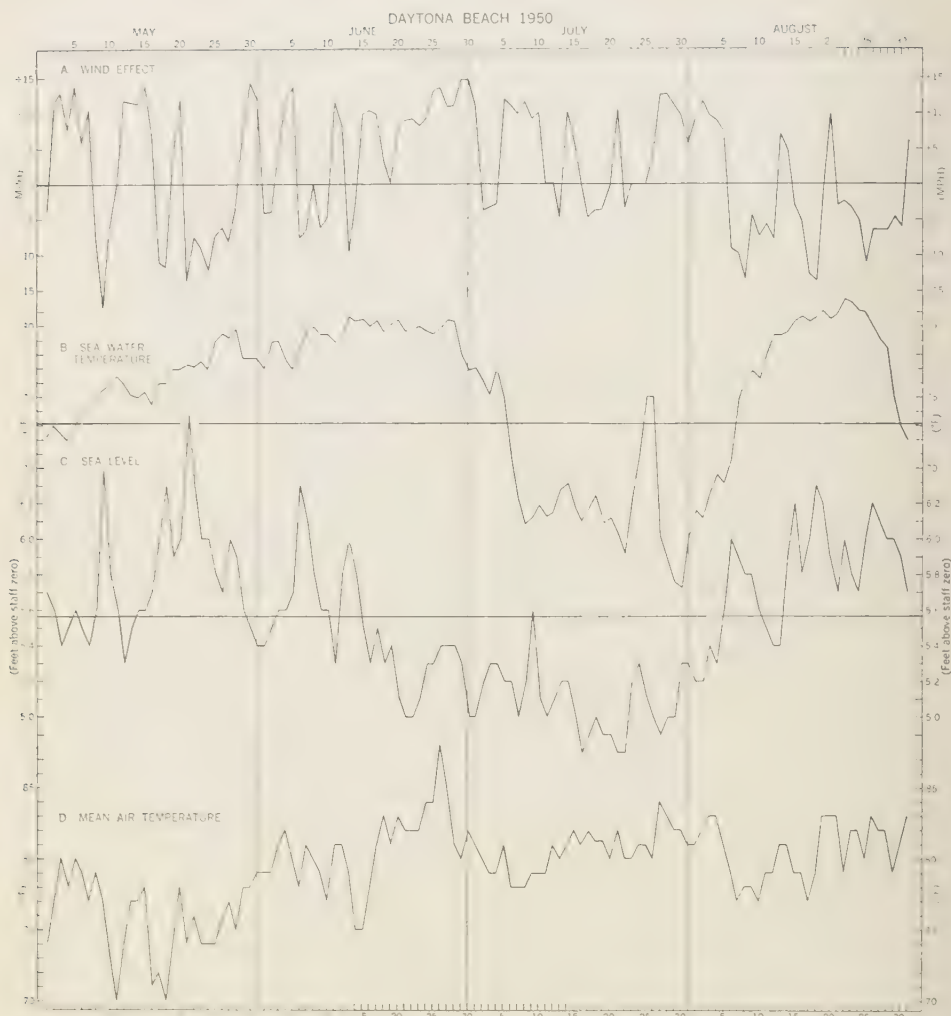


FIG. 6. Variations in (A) wind effect (see text for explanation), (B) sea-water temperature (horizontal line is five-year mean), (C) sea level (horizontal line is 19-year mean), and (D) mean air temperature, Daytona Beach, Florida, May 1 to September 1, 1950

the winds became negative and, except for four days, they remained negative until the end of the observations on September 1.

Like the wind effect, sea level (Fig. 6c) vacillated from May 1 until June 14 when the curve dropped below the long-term mean, and except for one day (July 9) it remained below the mean for over seven weeks until August 5. Daily sea level rose above the long-term mean

on August 5 and remained above it until the end of the observations with the exception of three days, again coinciding with a reversal in the wind effect. It is apparent, therefore, that during the period of observations local winds blowing for more than seven weeks in directions that would tend to move the surface water seaward were accompanied by a lowering of the daily sea level during those seven weeks in

amounts up to 0.8 ft. This good correlation between wind direction and sea level also indicates that any errors introduced through the use of prevailing wind and average velocity as mentioned above are not of sufficient magnitude to affect the results seriously.

Sea-water temperature (Fig. 6b) rose irregularly from May 1 until the start of the positive wind effect on June 13-14 when the rise terminated and the curve leveled off until June 29. On June 29-30 the water temperature dropped, the drop coinciding exactly with the time of the two highest values of positive wind effect (14.7 and 14.6 mi/hr). Temperatures remained below the five-year mean (horizontal line, Fig. 6b) for 31 days, except for a two-day period (July 25-26) which coincided with the end of a three-day period of zero wind effect and higher sea level. The period of water temperatures colder than the five-year mean terminated on August 7, the day following the return from positive to negative wind effect and the second day following the return of daily sea level to values above the long-term mean.

*Discussion*—The graphs in Figure 6 strongly support the upwelling hypothesis, for the period of colder water and lowered sea level both occurred during a protracted period of strong winds which tended to move surface water away from the coast, and all three factors returned to normal at approximately the same time (August 4-6). It is interesting that although the normal summer rise in temperature was halted by the advent of strong positive winds in mid-June, it was only after two weeks of strong steady positive winds that a large temperature decrease was noted. This indicates either that the actual upwelling had not started until then or that the upwelling had started earlier but had brought up only isothermal water for the first two weeks. The simultaneous termination of the positive wind, negative sea level, and colder water temperature suggests that the upwelling mechanism developed no momentum but stopped, at least at the coast, as soon as the motivating force was terminated.

Wagner [1957b] has shown that in summer off Miami the thermocline is much shallower (at about 150 ft) than in winter (at about 400 ft). Assuming that the area off Daytona Beach also has a shallower thermocline in summer than in winter, then any offshore movement of surface

water would tend to be reflected by greater decreases in water temperature in summer than in winter. Wagner [1957a] has proposed that the cooler water temperatures at Daytona Beach are related to tidal mixing and bottom configuration during times of shallow thermocline, a mechanism comparable to that used by Stevenson and Gorsline [1956] to explain the "cold spots" found off San Pedro, California. However, the present writers feel that the restriction of the colder water to one area, its usual association with lowered sea level, and its occurrence only with periods of offshore winds strongly supports upwelling rather than tidal mixing as the dominant cause.

Densities observed during the upwelling period showed a marked increase; however, when these values were expressed in terms of salinity at standard temperature, there was little variation other than a slight lowering of the salinity following heavy rains (2.36 inches on June 28) in late June and early July. This indicates that the higher densities observed during the upwelling period were due primarily to the lower temperature and suggests that any vertical circulation was in essentially isohaline water.

Variations in average daily air temperature recorded at Daytona Beach are shown graphically (Fig. 6d) along with the other variables. Abrupt changes in air temperature coincide almost exactly with changes from positive to negative wind effect. For example, during the period from May 1 to June 14 warmer air temperatures are found with the positive winds of May 5, 15, and 20 and June 4, colder air temperatures accompany the negative winds of May 10, 18, and 24 and June 10. After the upwelling had started to bring in colder water (June 29), air temperatures continued to reflect changes in the wind direction; however, air temperatures were generally below normal during that period. Unlike the other variables, the air temperature continued below normal throughout the remainder of the observation period, a fact which suggests that although colder water was no longer present at the coastal stations, it remained offshore and continued to cool the air returned to the coast as onshore winds.

#### REFERENCES

- DEFANT, A., Theoretische Überlegungen zum Phänomen des Windstaus und des Auftriebes



- an ozeanischen Küsten, *Deutsch. Hydrogr. Zs.*, **5**, 69-80, 1952.
- DEVEAUX, E. S., Meteorological trend and the apparent rise in sea level along the South Carolina coast, *Mon. Wea. Rev.*, **83**, 217-224, 1955.
- EKMAN, V. W., On the influence of the Earth's rotation on ocean currents, *Ark. Mat. Ast. Fysik.*, **2**, 52 pp., 1905.
- GREEN, C. K., Summer upwelling, northeast coast of Florida, *Science*, **100**, 546-547, 1944.
- HIDAKA, KOJI, A contribution to the theory of upwelling and coastal currents, *Trans. Amer. Geophys. Union*, **35**, 431-444, 1954.
- LONGARD, J. R., AND R. E. BANKS, Wind-induced vertical movement of the water on an open coast, *Trans. Amer. Geophys. Union*, **33**, 377-380, 1952.
- MARMER, H. A., Tidal datum planes, *USC&GS Spec. Pub. 135*, rev. 1951, U. S. Govt. Print. Off., 142 pp., 1951.
- MILLER, A. R., The effect of steady winds on sea level at Atlantic City, *Met. Monogr.*, **2**, 24-31, 1954.
- MILLER, A. R., The effects of winds on water levels on the New England coast, *Limnol. Oceanog.*, **3**, 1-14, 1958.
- PATTULLO, JUNE, W. MUNK, R. REVELLE, AND E. STRONG, The seasonal oscillation in sea level, *J. Mar. Res.*, **14**, 88-155, 1955.
- STEVENSON, R. E., AND D. S. GORSLINE, A shoreward movement of cool subsurface water, *Trans. Amer. Geophys. Union*, **37**, 553-557, 1956.
- SVERDRUP, H. U., On the process of upwelling, *J. Mar. Res.*, **1**, 155-164, 1938.
- SVERDRUP, H. U., AND R. H. FLEMING, The waters off the coast of Southern California, March to July 1937, *Bul. Scripps, Inst. Oceanog.*, **4**, 261-378, 1941.
- U. S. COAST AND GEODETIC SURVEY, Surface water temperatures at tide stations, Atlantic Coast, North and South America, *USC&GS Spec. Pub. 278*, U. S. Govt. Print. Off., 69 pp., 1955.
- U. S. COAST AND GEODETIC SURVEY, Density of sea water at tide stations, Atlantic Coast, North and South America, *USC&GS Pub. 31-2*, U. S. Govt. Print. Off., 72 pp., 1957.
- U. S. WEATHER BUREAU, *Normal surface wind data for the United States*, U. S. Govt. Print. Off., 1942.
- WAGNER, L. P., Note on the Daytona Beach cold water, *Tech. Rep. Marine Lab., Univ. of Miami*, ref. 57-9, pp. 81-82, 1957a.
- WAGNER, L. P., The general seasonal temperature pattern of the section between Miami and Bimini, *Tech. Rep. Marine Lab., Univ. of Miami*, ref. 57-9, pp. 42-43, 1957b.
- (Communicated manuscript received April 7, 1958, and, as revised, May 26, 1958.)

## Tracing Beach Sand Movement with Irradiated Quartz

D. L. INMAN AND T. K. CHAMBERLAIN\*

*Scripps Institution of Oceanography  
La Jolla, California*

**Abstract**—The mechanics of transportation of sand under the influence of wave action was studied using artificially induced radioactivity. A sample of naturally occurring quartz sand is taken from the area of study, subjected to slow neutron irradiation, and then reintroduced into the area. The movement of the quartz is traced by sampling the sediment on the bottom and assaying for phosphorus-32, which is the principal radio isotope formed from the slow neutron irradiation of natural quartz sands.

In a field experiment, 860 g of irradiated quartz sand was reintroduced as a point source on the sea floor off Scripps Beach. The experiment showed that the dispersal of sand by wave action was more rapid than expected, and that the movement of this small amount of sand could be followed for about 7 to 24 hr. The irradiated sand was released on the bottom by divers and samples were obtained on a grid at the time intervals of  $\frac{1}{4}$ , 1,  $3\frac{1}{2}$ ,  $7\frac{1}{2}$ , and 24 hr, and several days later. At the point of introduction the water was 10 ft deep and the surface waves were about  $1\frac{1}{2}$  ft high. The waves caused the sand to spread in an elliptical pattern with an elongate onshore-offshore axis. The sampling indicated that the 860 g of irradiated fine sand had been dispersed over an area of approximately three-quarters of a square mile in 24 hr. The sensitivity of detection for this experiment was approximately 1 grain of sand in 100,000.

**Introduction**—Until recently there has been no satisfactory means of tagging grains of sand so that their movement could be studied in the beach and nearshore zone. The advent of readily available high flux neutron sources for activating sand samples has given impetus to studies involving the tracing of sediment movements. The purpose of this investigation was to perfect a field technique using low levels of radioactivity in sands, and to apply this technique in studying the rate of dispersion of sand under wave action near shore.

Previous work with radio isotopes as tracers of sand has principally been confined to the use of foreign particles such as ground glass, which are added to the sand under study. Also, previous investigators have employed levels of activity too high for use along California beaches [Hours and others, 1955; Inose and Shiraishi, 1956; Putman and Smith, 1956; Arlman and others, 1957; Davidsson, 1958]. Foreign particles are not usually suitable for beach sand tracers because their hydraulic behavior may be different from that of the natural sands. To be satisfactory, the natural tracer of sand movements should be: (1) not a health hazard, (2) of the

same size, density, and shape as one of the major components under study, (3) easily and rapidly distinguishable from the sand mass, (4) inexpensive and available in relatively large amounts, and (5) able to retain its distinguishing properties over times comparable to the times of the fluid processes. Thus, the problem was to seek out a tracer that would fulfill these requirements and to develop field techniques that could be used in the difficult conditions encountered near the surf zone.

Since quartz is one of the dominant components of most sands, a method of tagging it was highly desirable. The technique by which Goldberg and Inman [1955] utilized phosphorus-32, which is formed when normally-occurring phosphorus in quartz sand is subjected to slow neutron irradiation, satisfactorily fulfills the requirements for a tracer. In addition phosphorus-32 does not constitute a health hazard when used in this manner near populated beaches. Phosphorus is a low level beta emitter which occurs as a minor impurity in the order of hundreds of parts per million in quartz. The phosphorus mineral involved may be an orthophosphate, possibly berillinite ( $\text{AlPO}_4$ ), which is isostructural with quartz. Since each radioactive quartz particle can be detected on photosensitive film, sensi-

\* Contribution from Scripps Institution of Oceanography, New Series.

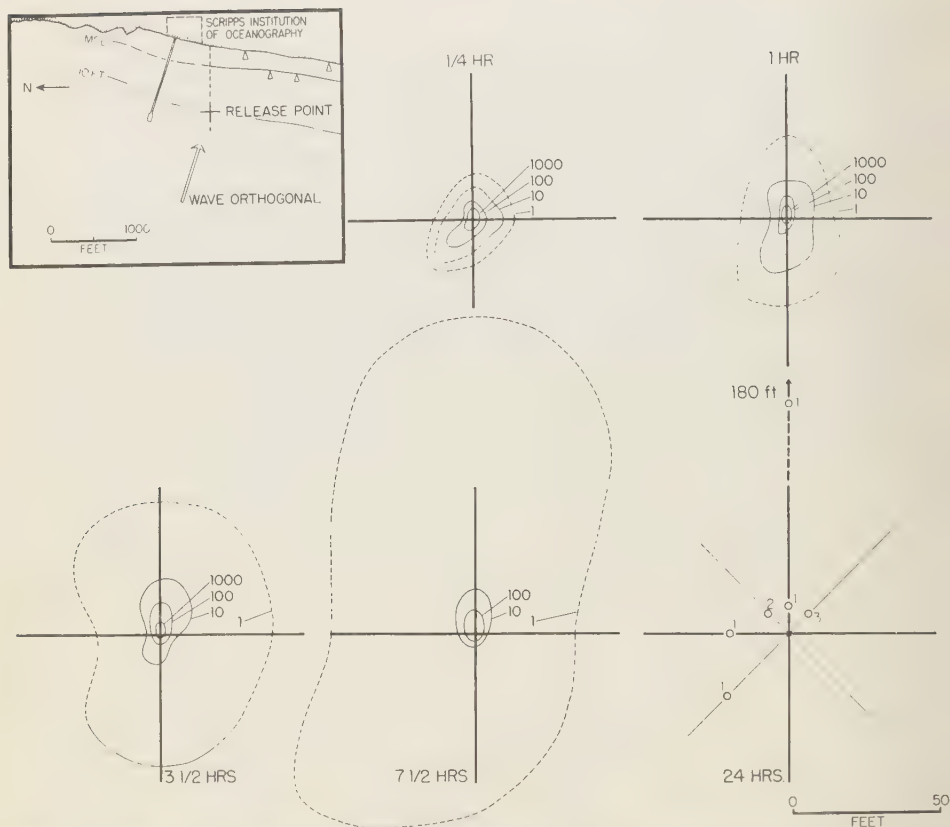


FIG. 1—Distribution of irradiated sand particles with time. The contours are lines of equal concentration of surface particles, and their value is expressed in number of irradiated particles per 0.1 sq ft (100 cm<sup>2</sup>). At 24 hr only six offshore samples gave returns, which are shown as individual points. The four positive beach counts for the 24-hr sampling are plotted in the inset. Orientation of grid is shown in the inset.

tivity of the assay is limited only by size of sample and size of film. Single irradiated quartz particles can easily be detected when mixed with about one million inactive particles. Phosphorus-32 has a half-life of 14.6 days, and irradiated quartz particles could be detected for several months.

To carry out the study, a natural sample of sand was obtained from the environment to be studied. The quartz grains were concentrated by partial removal of other constituents, the resulting concentrate was irradiated, and it was then reintroduced into the natural environment by swimmers equipped with self-contained underwater breathing apparatus (SCUBA). After reintroduction of the irradiated sand on the

bottom, the area was periodically sampled to determine the rate of dispersion of irradiated particles. Two types of samples were taken: (1) the surface layer of sand, obtained by impressing a grease-coated 3 × 5-in card on the bottom, and (2) cores, obtained by forcing small plastic vials into the bottom. The activity of the surface sample was assayed by laying the sample card on photosensitive film and, after exposure and development, counting the exposed spots. The cores were solidified with plastic resin and sectioned before assaying.

Since it was desirable to use an area that was convenient to reach from the laboratory and one in which good wave and tidal data could be obtained, the sandy beach and shelf immediately

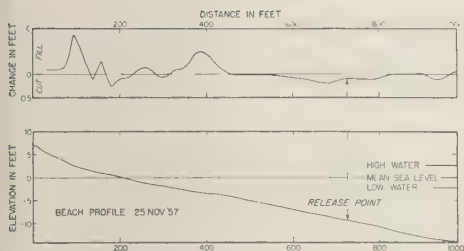


FIG. 2—Changes in sand level along the beach profile during first 40 hr of the experiment. Erosion (cut) occurred in the vicinity of the release point; deposition (fill) occurred on the beach. Location of profile shown in Figure 1.

adjacent to the Scripps Institution of Oceanography was chosen as the field locality in which to carry out the experiments (Fig. 1). The shelf is flat and is uniformly covered with fine and very fine sand. The irradiated sand was released on the bottom by swimmers at a point where the water depth was about 10 ft. During the observations the surface waves were about  $1\frac{1}{4}$  ft high and the wave period was about 14 sec (Table 1).

**Laboratory method**—The natural sand collected from the experiment site was composed of about 68 pct quartz, 20 pct feldspar, 10 pct heavy minerals, and 2 pct shell fragments [Inman, 1953]. Thus it was necessary to process the sample in order to concentrate the quartz. The sample was first washed and treated with dilute

TABLE 1—Wave and tide data

Time after release	Depth of water at release point	Wave <sup>a</sup> height	Wave <sup>a</sup> period
hr	ft	ft	sec
$\frac{1}{4}$	11.3	1.3	15.0
1	11.6	1.2	14.3
$2\frac{1}{2}$	11.9 (highest water)		
$3\frac{1}{2}$	11.7	1.1	13.9
$7\frac{1}{2}$	8.6	1.4	13.3
10	7.7 (lowest water)		
24	11.0	1.9	13.0

<sup>a</sup> Height and period of the highest one-third of the waves measured during a 20-min interval. The maximum wave measured during each interval was about 1.6 times greater.

HCl to remove the calcareous shell material, then dried and sieved, and the smallest size group discarded because it contained little quartz. The magnetic minerals were removed with an electromagnetic separator, after which the sample was placed in hot *aqua regia* for 24 hours. The sample resulting from the above treatment was nearly 98 pct quartz and feldspars and compared closely in size distribution with the natural sand of the experimental area (Fig. 3).

The concentrate of 860 g of sand, prepared as outlined above, was sent for irradiation to the Union Carbide Nuclear Company, Oak Ridge National Laboratory, Oak Ridge, Tennessee. There it was placed in the atomic pile and subjected to slow neutron irradiation with a nominal flux of  $10^{12}$  neutrons/cm<sup>2</sup> sec for 14 days. Following irradiation and prior to shipping, the sample was cooled for seven days to allow for the removal of the induced Na and K activities in the feldspar impurities. After receipt at the laboratory, the irradiated sand measured approximately 16 mr/hr at the surface of the container. Beta absorption measurements were taken, and

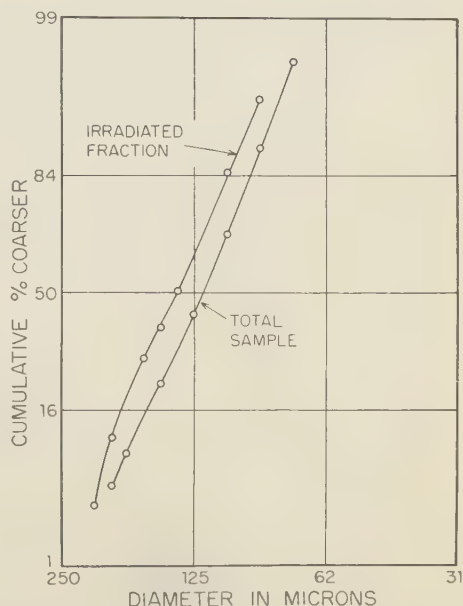


FIG. 3—Particle size distribution curves for the total sample of natural sand and the concentrate of quartz sand which was irradiated and reintroduced into the area of investigation.



they showed the presence of the phosphorus 1.7 Mev beta particle. Measurement on half-life indicated that the short life contaminants did not contribute significantly to the over-all activity.

The sample was dyed an intense blue (Night Blue, National Aniline No. 598) before introduction to the experiment site. Coloring the tracer permitted the first stages of the dispersion to be followed visually.

**Counting technique**—To determine the quantity of irradiated sand in each sample collected, the  $3 \times 5$ -in sampling cards, covered with sand grains, were brought from the field and wrapped face to face with No-Screen X-ray film in a light-tight package. The film was exposed 24 hr in this manner, then developed and fixed. The irradiated sand grains showed themselves quite easily as black spots on the resulting transparent film negative.

The small cores collected in the field were brought into the laboratory and solidified by the injection of plastic resin. They were then sectioned lengthwise and laid upon photosensitive film for the required exposure period. The result was similar to that obtained from the surface samples, and counts could easily be made of the irradiated grains.

**Sensitivity**—Tests were conducted on a portion

of the irradiated sample to determine the sensitivity of the photosensitive film in recording individual radioactive sand grains and to determine the most desirable exposure time to use at various decay times. It was found that thin wrapping paper placed between the irradiated particles and the X-ray film absorbed most of the radiation, but that "Saran Wrap", an extremely thin sheeting of vinylidene chloride made by Dow Chemical Company, allowed the radiation to penetrate and register on the film. Saran Wrap was used to protect the film from the greased sand sample in all experiments. Thirty days after irradiation an exposure time of 4 hr gave weak returns from the control sample, while 12 hr was adequate to give strong signatures from all irradiated grains. Four months after irradiation, high contrast could still be obtained from all particles in the control sample with an exposure time of 16 hr. Exposure times longer than 48 hr were excessive and resulted in clouding the film, thereby obscuring particle counts.

Prior to the release of irradiated sand, 32 samples of sand were collected from the area of investigation to determine the natural background activity of the region. These were processed in the usual way with an exposure time of 24 hr. Although each sample contained about 100,000 particles of natural sand, no

TABLE 2—Distribution of irradiated sand particles ( $2.44 \times 10^8$  irradiated particles were released)<sup>b</sup>

Time after release	Number of surface samples	Number of cores	Number of particles accounted for by sampling			Area of <sup>c</sup> distribution,
			Surface	Subsurface	Total	
hr $\frac{1}{4}$	32	0	$3.25 \times 10^6$	$1.19 \times 10^6$	$1.22 \times 10^6$	sq ft $0.73 \times 10^3$
1	37	9	$0.93 \times 10^6$	$1.19 \times 10^6$	$1.19 \times 10^6$	$1.5 \times 10^3$
$3\frac{1}{2}$	42	5	$0.19 \times 10^6$	$0.77 \times 10^6$	$0.79 \times 10^6$	$5.3 \times 10^3$
$7\frac{1}{2}$	42	11	$0.24 \times 10^6$	$0.31 \times 10^6$	$0.31 \times 10^6$	$17.2 \times 10^3$
24	74	5	scattered returns approximately			$24 \times 10^{6d}$
7 to 23 days	84	9	no positive returns			indeterminate

<sup>b</sup> 860 g of very fine sand were released.

<sup>c</sup> Area over which concentration of irradiated particles per sq ft was 10 or more. This is a surface concentration of 1 in  $10^5$  particles.

<sup>d</sup> Assuming total number of irradiated particles to be dispersed so that concentration per sq ft is 10 or less as indicated by sampling.

positive exposures were found. Supporting this assumption of no detectable background activity were the results of the actual sampling itself. The last irradiated grains were detected from the sampling at 24 hr. Subsequent to this, 84 surface samples were taken in the area, none of which showed the presence of irradiation.

*Field procedure*—The field work was carried out by SCUBA-equipped swimmers operating from a small boat. Concurrent with the sand sampling a pressure-type wave meter was used to record wave conditions, and current measurements were taken using a captive drogue. Principal operations were concluded in 24 hr, although background sampling was continued for 23 days. A total of 311 surface samples and 39 cores was collected and processed (Table 2).

The irradiated sand was carried to the center of the grid in a plastic bottle by a swimmer. There the plastic bottle was opened and the irradiated sand poured out onto the bottom, covering an area of about 3 sq ft. Samples were taken at predetermined positions on the grid at  $\frac{1}{4}$ , 1,  $3\frac{1}{2}$ ,  $7\frac{1}{2}$ , and 24 hr, (Fig. 1, Table 2). Also, background samples were obtained throughout the area 7 and 23 days after the release, but they did not show the presence of irradiated particles.

A permanent sampling grid was established prior to the actual experiment. It consisted of a series of reference stakes driven into the bottom at 10-ft intervals along eight ranges radiating from the release point. The major onshore-offshore range extended through the breaker zone onto the beach (Fig. 1). The beach profile and its changes are shown in Fig. 2.

The principal sampling was performed with 3  $\times$  5-in paraffin-coated paper cards, covered on one side with Varcon #4-1645 Water Pump Grease. Each sample was taken by pressing a sampling card onto the bottom so that the surface layer of sand grains adhered to the coating of grease. In addition, subsurface samples were obtained by forcing small plastic vials into the bottom. The plastic vials, approximately one inch in diameter by three inches long, were quite efficient in obtaining relatively undisturbed sediment samples.

The distribution patterns of irradiated grains showed extremely rapid rates of dispersion; greater than were expected under such low wave conditions. The transport rate for some particles

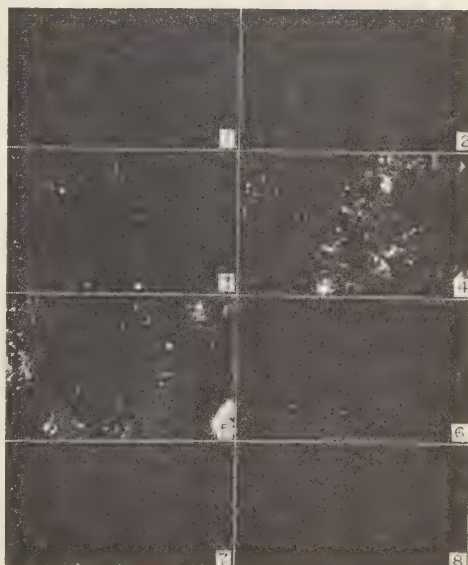


FIG. 4—Eight representative surface samples taken along the onshore-off-shore range line of Figure 2. Samples were taken  $\frac{1}{4}$  hr after the release of irradiated material. These are prints from photo-sensitive film negatives used in counting the irradiated grains. Each white dot represents one irradiated sand grain. The location of the samples relative to the release point is as follows: (1) 10 ft shoreward, (2)  $7\frac{1}{2}$  ft shoreward, (3) 5 ft shoreward, (4)  $2\frac{1}{2}$  ft shoreward, (5)  $2\frac{1}{2}$  ft seaward, (6) 5 ft seaward, (7)  $7\frac{1}{2}$  ft seaward, (8) 10 ft seaward.

was as high as 1.5 ft/min during the first  $\frac{1}{4}$ -hr period. The initial distribution pattern of surface counts showed dispersion in all directions, with an elongate offshore-onshore axis. In 7 to 24 hr the net onshore transport of sand overshadowed the lateral and offshore dispersion, and the entire surface distribution pattern indicated an onshore migration. Of 26 surface samples collected from the beach at 24 hr, four showed single irradiated particles. The locations of these samples are shown by triangular symbols on the inset of Fig. 1. Both the surface and subsurface samples showed an increase in irradiated grains with time in a shoreward direction, and for the most part the highest counts were found shoreward of the release point at all sampling times.

The rapid rate of dispersal of sand observed for these relatively quiescent wave conditions may in part result from suspension associated

with the alternate erosion and steepening of the 3 to 4-in sand ripples which formed on the bottom. Steep ripples with a height to length ratio of 1:6 were observed to form during the gentle surges accompanying the smaller waves. During the stronger surges accompanying the occasional largest waves (1 to 2-ft in height), these steep ripples were eroded to a more gentle form with a steepness ratio of about 1:12. The first phase in the erosion of the ripples was evidenced by a plume or vortex of suspended sand which extended about six inches above the bottom. For slightly higher wave conditions the ripples are always gentle and thus do not give rise to this type of periodic suspension.

Accurate measurements of the change in sand level were obtained by comparing the length of reference rods exposed along the major axis of the sampling grid, following the procedure of *Inman and Rusnak* [1956]. The sand level measurements show that erosion occurred in the vicinity of the release point and that deposition of sand occurred on the beach and, to a lesser extent, offshore (Fig. 2).

The 39 cores taken in the sampling area showed the occurrence of vertical mixing of sand. The subsurface concentration at the release point decreased with time, while the depth of mixed sand increased. At 24 hr, irradiated sand was found as deep as 0.2 ft below the surface of the core taken at the release point. This burying may in part result from the disturbance of previous coring. Cores taken away from the release point showed a decrease in concentration of irradiated material following the one hour sampling, with the exception of cores taken in the onshore direction where a maximum concentration occurred 10 ft from the release point at  $7\frac{1}{2}$  hr. A maximum depth of mixing of 0.06 ft was observed at this time.

*Summary and conclusions*—1. The procedure used in this radioactive tracer experiment was: (a) collection of natural sand sample from the area under investigation, (b) processing sample to concentrate quartz grains, (c) irradiation of processed sample, (d) reintroduction of irradiated sample as a point source on the bottom by SCUBA-equipped swimmers, (e) collection and assay of samples for radioactive particles by radioautography. The irradiated sand did not constitute an apparent health hazard.

2. The sensitivity for this experiment was one

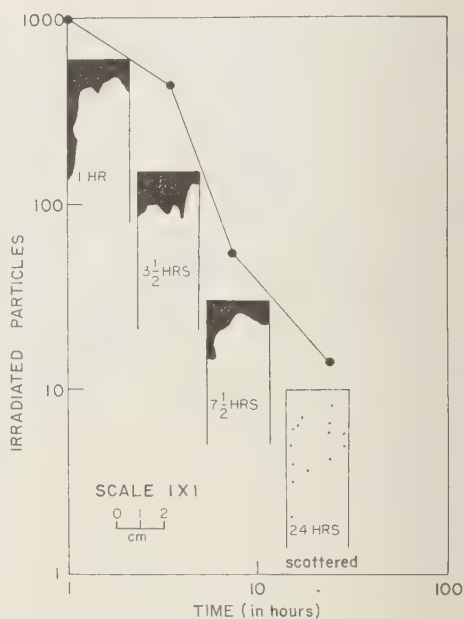


FIG. 5—Concentration of irradiated particles in cores taken at the release point. Vertical distribution of irradiation is shown as black on the diagram of each core. Cores were lithified, sectioned, then exposed to photosensitive film. The extension of the zone of irradiated sand down along the side of the coring tube is the result of coring method.

particle per 100,000 particles. However, greater sensitivity is practical and is limited only by size of photosensitive film and size of sample. Sensitivity tests showed that irradiated particles could be detected for periods of several months.

3. The field experiment indicated a rapid rate of dispersion of sand from the point of release, where the waves were about  $1\frac{1}{4}$  ft high and the water depth was about 10 ft. A net sand transportation rate of  $1\frac{1}{2}$  ft/min was measured near the release point.

4. Dispersion of sand was most rapid in the onshore and offshore directions. Irradiated grains were detected on the beach approximately  $\frac{1}{4}$  mi from the release point in 24 hr.

5. Sampling indicated a net onshore transportation of irradiated particles, and that the particles were dispersed over an area as great as 0.75 sq mi in 24 hr (Table 2).

6. Cores near the release point showed

irradiated particles as much as 0.06 ft below the surface of the sand.

7. Measurements of sand level change show that a net erosion of about 0.1 ft occurred in the vicinity of the release point, while deposition of sand occurred on the beach foreshore (Fig. 2). It seems probable that the rate of dispersion of sand would be quite different for irradiated material released in a zone of net deposition.

*Acknowledgments*—This study was sponsored by the Office of Naval Research under contract with the University of California, and by the Water Resources Center of the University of California.

The field phase of the investigation was performed principally by Earl Murray. Assistance in the field work and analysis of data was given by Helene Flanders. Valuable suggestions and guidance during the course of the study were contributed by E. D. Goldberg, R. S. Arthur, and J. D. Frautschy.

#### REFERENCES

- ARLMAN, J. J., P. SANTEMA, AND J. N. SVASEK, Movement of bottom sediment in coastal waters by currents and waves; measurements with the aid of radioactive tracers in the Netherlands, *Deltadienst, Rijkswaterstaat, Ministry of Transport and Waterstaat, The Netherlands, Progress Report*, 1-56, June 1957.
- DAVIDSSON, JAN, Investigations of sand movement using radio-active sand, *Lund Studies in Geography Ser., A Physical Geography*, 12, 107-126, 1958.
- GOLDBERG, E. D. AND D. L. INMAN, Neutron-irradiated quartz as a tracer of sand movements, *Bulletin of the Geological Society of America*, 66, 611-13, 1955.
- HOURS, R., W. D. NESTEROFF, AND V. ROMANOVSKY, Methode d'etude de l'evolution des plages par traceurs radio-actifs, *Centre de Recherches et D'etudes Oceanographiques*, 1, no. 11, 1-7, 1955.
- INMAN, D. L., Areal and seasonal variations in beach and nearshore sediments at La Jolla, California, *Beach Erosion Board, Corps of Engineers*, Tech. Memo., no. 39, 1-34, 1953.
- INMAN, D. L. AND G. A. RUSNAK, Changes in sand level on the beach and shelf at La Jolla, California, *Beach Erosion Board, Corps of Engineers*, Tech. Memo., no. 79, 1-72, 1956.
- INOSE, S. AND N. SHIRAISHI, The measurement of littoral drift by radio-isotopes, *The Dock and Harbour Authority, Hokkaido Development Bureau, Japan*, 284-88, January, 1956.
- PUTMAN, J. L. AND D. B. SMITH, Radioactive tracer techniques for sand and silt movements under water, *International Journal of Applied Radiation and Isotopes*, 1, 24-32, 1956.

(Manuscript received August 29, 1958)





## Rapid Gravity Computations for Two-Dimensional Bodies with Application to the Mendocino Submarine Fracture Zone\*

MANIK TALWANI, J. LAMAR WORZEL, AND MARK LANDISMAN

*Lamont Geological Observatory,  
(Columbia University)  
Palisades, New York*

**Abstract**—Expressions are derived for the vertical and horizontal components of the gravitational attraction due to a two-dimensional body of arbitrary shape by approximating it to an  $n$ -sided polygon. These expressions are put in forms suitable for solution by a high-speed digital computer. As an example of the application of this method, the crustal section across the Mendocino fracture zone is deduced from the gravity anomalies. Assuming the crust to consist of a single homogeneous layer, overlain by water and sediment, it is found to be about three km thicker to the north of the fracture zone than to the south of it.

Many geological structures are approximately linear, and the problems connected with them can be solved with two-dimensional forms of analysis. For gravity computations, Nettleton [1940] has determined the criteria for making an adequate two-dimensional computation. Various methods exist for the computation of the gravitational attraction caused by irregularly shaped two-dimensional bodies. These methods can be divided into two categories. In the first category are those that involve the use of gratitudes, dot charts, or other such graphical computing aids. While in theory these methods can be made as precise as one pleases, merely by increasing the scale to which the graticule is constructed, in actual practice this may be difficult, if not impossible.

In the second category lie those methods that involve breaking up the irregularly shaped bodies into several smaller bodies of different sizes but of shapes that are regular and for which the gravitational attraction can be easily computed. A convenient form of regular body to use is the rectangular block, as proposed by Vening Meinesz and others [1934]. Here again the method can be made as precise as one pleases by using a sufficiently large number of small blocks [see, for example, Shurbet and others, 1956]. However, the computations become increasingly more tedious as the number of blocks is increased. Further, the blocks may be so small that their

individual contributions at distant points are neglected even though the total sum of their small contributions is appreciable. This may cause considerable error in computations.

The periphery of any two-dimensional body can be approximated closely by a polygon, by making the number of sides of this polygon sufficiently large. Analytical expressions can be obtained for both the vertical and horizontal components of the gravitational attraction due to this polygon at any given point. These expressions, then, can be used without any limitations to the size or position of the body. The present method involves the use of these expressions. The accuracy depends only on how closely the polygon fits the given body, and can be increased by increasing the number of sides of the polygon. It will be recognized that an irregularly shaped two-dimensional body can be more easily approximated by a polygon than by rectangular blocks. The computations involved in solving the expressions to obtain the components of gravitational attraction are lengthy and tedious, but being iterative are readily programmed for solution by a digital computer. A program for use with the IBM 650 has been made, and the machine time required for obtaining both the vertical and horizontal components of gravitational attraction for an  $n$ -sided polygon at a single point is approximately equal to  $2.5n$  seconds.

Let  $ABCDEF$  (Fig. 1) be a given polygon with  $n$  sides and let  $P$  be the point at which the

\* Lamont Geological Observatory Contribution No. 318

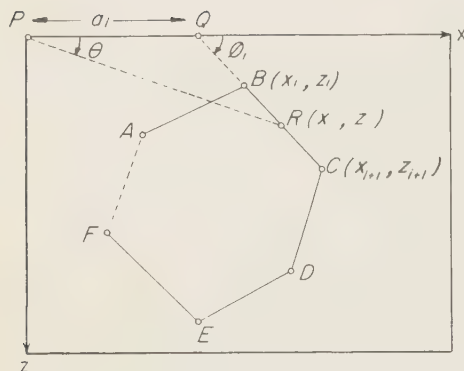


FIG. 1—Geometrical elements involved in the gravitational attraction of an  $n$ -sided polygon

attraction due to this polygon has to be determined. Imagine  $P$  to be the origin of an  $xz$  system of coordinates, where the polygon also lies in the  $xz$  plane. Let  $z$  be defined positive downwards (vertical) and let  $\theta$  be measured from the positive  $x$  axis towards the positive  $z$  axis.

It has been shown by Hubbert [1948] that the vertical component of gravitational attraction due to such a two-dimensional body is, at the origin, equal to

$$2G\rho \oint z d\theta,$$

the line integral being taken along its periphery, where  $G$  is the universal constant of gravitation and  $\rho$  is the volume density of the body. It can be shown by a method similar to Hubbert's that the corresponding expression for the horizontal component of gravitational attraction is given by  $2G\rho \oint x d\theta$ .

Let us evaluate the two integrals  $\oint z d\theta$  and  $\oint x d\theta$  for the above polygon. The contribution to  $\oint z d\theta$  from, say, the side  $BC$  of the polygon can be first computed. Produce  $CB$  to meet the  $x$  axis at  $Q$  at an angle  $\phi_i$ . Let  $PQ = a_i$ . Now

$$z = x \tan \theta \quad (1)$$

for any arbitrary point  $R$  on  $BC$ . Also

$$z = (x - a_i) \tan \phi_i \quad (2)$$

From (1) and (2)

$$z = \frac{a_i \tan \theta \tan \phi_i}{\tan \phi_i - \tan \theta},$$

or

$$\int_{BC} z d\theta = \int_B^C \frac{a_i \tan \theta \tan \phi_i}{\tan \phi_i - \tan \theta} d\theta \equiv Z_i.$$

Similarly it can be shown that

$$\int_{BC} x d\theta = \int_B^C \frac{a_i \tan \phi_i}{\tan \phi_i - \tan \theta} d\theta \equiv X_i.$$

The vertical component of gravitational attraction  $V$  and the horizontal component  $H$ , due to the whole polygon, are then given respectively by

$$V = 2G\rho \sum_{i=1}^n Z_i$$

and

$$H = 2G\rho \sum_{i=1}^n X_i,$$

the summations being made over the  $n$  sides of the polygon.

It now remains to solve the integrals involved in the expressions for  $Z_i$  and  $X_i$ .

In the most general case it can be shown that

$$\begin{aligned} Z_i &= a_i \sin \phi_i \cos \phi_i \left[ \theta_i - \theta_{i+1} \right. \\ &\quad \left. + \tan \phi_i \log_e \frac{\cos \theta_i (\tan \theta_i - \tan \phi_i)}{\cos \theta_{i+1} (\tan \theta_{i+1} - \tan \phi_i)} \right] \\ X_i &= a_i \sin \phi_i \cos \phi_i \left[ \tan \phi_i (\theta_{i+1} - \theta_i) \right. \\ &\quad \left. + \log_e \frac{\cos \theta_i (\tan \theta_i - \tan \phi_i)}{\cos \theta_{i+1} (\tan \theta_{i+1} - \tan \phi_i)} \right] \end{aligned}$$

where

$$\theta_i = \tan^{-1} \frac{z_i}{x_i},$$

$$\phi_i = \tan^{-1} \frac{z_{i+1} - z_i}{x_{i+1} - x_i},$$

$$\theta_{i+1} = \tan^{-1} \frac{z_{i+1}}{x_{i+1}},$$

and

$$a_i = x_{i+1} + \frac{z_{i+1} - z_i}{z_i - z_{i+1}}.$$

Also see Figure 1.

The expressions for  $Z_i$  and  $X_i$  reduce to simple expressions in the following cases:

Case A—If  $x_i = 0$

$$Z_i = -a_i \sin \phi_i \cos \phi_i \left[ \theta_{i+1} - \frac{\pi}{2} \right. \\ \left. + \tan \phi_i \log_e \{ \cos \theta_{i+1} (\tan \theta_{i+1} - \tan \phi_i) \} \right] \\ X_i = a_i \sin \phi_i \cos \phi_i \left[ \tan \phi_i \left( \theta_{i+1} - \frac{\pi}{2} \right) \right. \\ \left. - \log_e \{ \cos \theta_{i+1} (\tan \theta_{i+1} - \tan \phi_i) \} \right].$$

Case B—If  $x_{i+1} = 0$

$$Z_i = a_i \sin \phi_i \cos \phi_i \left[ \theta_i - \frac{\pi}{2} \right. \\ \left. + \tan \phi_i \log_e \{ \cos \theta_i (\tan \theta_i - \tan \phi_i) \} \right] \\ X_i = -a_i \sin \phi_i \cos \phi_i \left[ \tan \phi_i \left( \theta_i - \frac{\pi}{2} \right) \right. \\ \left. - \log_e \{ \cos \theta_i (\tan \theta_i - \tan \phi_i) \} \right].$$

Case C—If  $z_i = z_{i+1}$

$$Z_i = z_i (\theta_{i+1} - \theta_i) \\ X_i = z_i \log_e \frac{\sin \theta_{i+1}}{\sin \theta_i}.$$

Case D—If  $x_i = x_{i+1}$

$$Z_i = x_i \log_e \frac{\cos \theta_i}{\cos \theta_{i+1}} \\ X_i = x_i (\theta_{i+1} - \theta_i).$$

Case E—If  $\theta_i = \theta_{i+1}$

$$Z_i = 0$$

$$X_i = 0.$$

Case F—If  $x_i = z_i = 0$

$$Z_i = 0$$

$$X_i = 0.$$

Case G—If  $x_{i+1} = z_{i+1} = 0$

$$Z_i = 0$$

$$X_i = 0.$$

Noting that  $\theta_i$ ,  $\theta_{i+1}$ ,  $\phi_i$ , and  $a_i$  can all be explicitly expressed in terms of the  $x_i$ 's and the

$z_i$ 's, we are able to obtain expressions for both  $V$  and  $H$  solely in terms of the  $x_i$ 's and the  $z_i$ 's. This is especially advantageous, since one of the simplest ways of defining the periphery of a body is to specify the coordinates of adjacent points at the vertices of the body. These are the coordinates  $x_i$ 's and  $z_i$ 's used in the computation. In addition, of course, it is necessary to specify the density of the body and the position of the points at which the attraction is to be calculated.

This method is illustrated by the following example in which the crustal structure along a profile in the Pacific crossing the Mendocino fracture zone has been deduced.

The Mendocino fracture zone has been described by *Menard and Dietz* [1952] and by *Menard* [1955]. The map in Figure 2 shows the position of the Mendocino escarpment and the submarine gravity stations in the area made on USS *Redfish* in 1952 [*Worzel and others*, 1955] and on the USS *Rasher* and USS *Raton* in 1954 [*Harrison and others*, 1957]. The gravity values at the *Redfish* stations and at those of the Harrison stations that were used in this study are listed in Table 1.

TABLE 1—Gravity stations<sup>a</sup>

Cruise	Sta.	Latitude N		Longitude W		Corrected sounding	Free-air anomaly
		°	'	°	'	m	mgal
<i>Redfish</i> (first cruise)	12	38	15.0	126	31.0	2503	-29
	13	38	38.8	127	11.3	4592	-44
	14	39	02.4	127	53.4	4396	-17
	15	39	53.9	129	28.4	4611	-46
	16	40	18.4	130	16.1	3635	-38
	17	40	42.3	131	02.1	3767	-12
	18	41	04.7	131	44.5	3754	-15
	19	41	30.8	132	35.3	3862	-19
	20	42	14.6	134	02.0	3887	-22
	21	42	36.9	134	47.7	3862	-19
	146	41	40	128	58	3352 <sup>b</sup>	+ 1
	147	40	44	128	14	3541 <sup>c</sup>	+11
	148	39	46	127	42	4551 <sup>c</sup>	-44
	149	38	58	127	16	4398 <sup>c</sup>	-19
<i>Harrison</i> and others [1957]	61	40	45.7	128	57.2	3330	+13
	62	40	34.7	128	57.2	3319	+19
	63	40	26.0	129	02.7	3319	+ 4
	64	40	12.7	129	01.3	4638	-79
	65	40	01.8	129	04.5	4738	-63
	66	39	49.2	128	58.8	4738	-45
	67	39	31.6	128	58.8	4484	-17

<sup>a</sup> For other stations in this area see Harrison and others [1957].

<sup>b</sup> Based on soundings from U.S.C.G.S. Chart 9000.

<sup>c</sup> Based on soundings from U.S.C.G.S. Chart 5002.



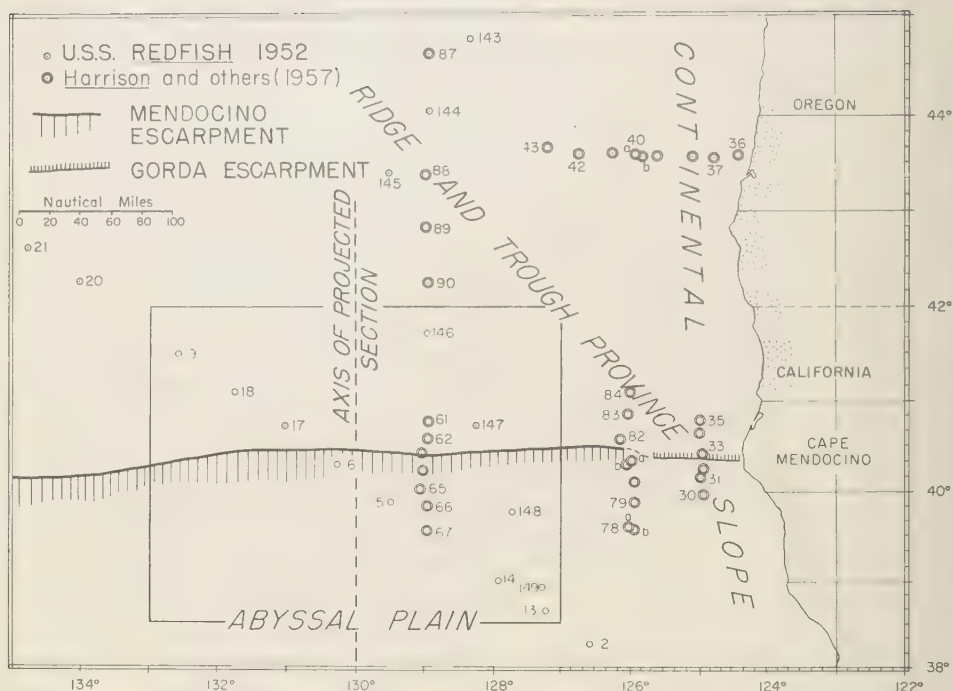


FIG. 2—Position of Mendocino and Gorda escarpments (modified from Menard and Dietz [1952]); gravity stations within the box are used in this study

While there are three *N-S* Harrison profiles in this area, only the westernmost one (Stations 61 to 67 and 85 to 90) can be considered a typical gravity profile across the Mendocino escarpment. Of the remaining two profiles, one crosses the Mendocino escarpment at its extreme eastern edge, and the other crosses the Gorda escarpment rather than the Mendocino escarpment (see Fig. 2). Even the first profile, north of the escarpment, soon passes into the "Ridge and Trough Province" [Menard and Dietz, 1951], a province of very rough topography. The situation is complicated by the fact that while the topography in this area seems to trend in an approximately *SW-NE* direction [Menard and Dietz, 1952], the tectonic trends as described by Heezen (personal communication) run in a *NW-SE* direction. A deduction of structure based on a two-dimensional gravity analysis, using gravity values obtained in such an area, would be inadequate.

However, some distance to the west of the Harrison profile, the main *Redfish* profile (Sta-

tions 12 to 21) lies north of the escarpment in comparatively smooth topography. But it runs in a *NW-SE* direction, crossing the *E-W* trending escarpment at an acute angle. This, again, is not very suitable for a two-dimensional gravity analysis. A compromise was effected by 'projecting' the *Redfish* gravity profile on to a *N-S* line ( $130^{\circ}\text{W}$ ). The topographic profile along this line was also made up from *Redfish* soundings. The minimum depth at the top of the escarpment is 3100 m (1700 fm), and the maximum depth at the foot of the escarpment is 4900 m (2700 fm). The average slope of the escarpment is about  $6^{\circ}$ . These figures are in reasonable agreement with those given by Menard and Dietz [1952] for their profiles along the latitudes ( $128^{\circ} 12'\text{W}$  and  $131^{\circ}\text{W}$ ). The Harrison stations and the *Redfish* stations from the profile that contains Stations 143 to 149 were also projected on to the same profile. However, the Harrison Stations 61 and 62 and the *Redfish* Stations 146 and 147 are at considerably different depths than those corresponding to their respective positions along the

profile. Therefore it was decided not to use these stations in making up the composite profile. They are discussed separately.

The composite profile was terminated on the north at 42°N. The main *Redfish* profile north of here diverges considerably from the line of projection, and the Harrison profile, as mentioned earlier, enters the very rough Ridge and Trough Province. Stations east of 127°W were also not used in the composite profile. It was felt that the proximity of the continental edge significantly affected the gravity values at these stations.

In the following calculations, the crust is assumed to consist of the water layer underlain by a single homogeneous layer of density 2.84 gm/cc, but of variable thickness. Beneath the crust, the mantle is assumed to have a density of 3.27 gm/cc. The crust is taken as the layer bounded at the bottom by the Mohorovicic discontinuity. The problem, then, is to compute the attraction of the water layer and subtract this from the observed anomalies. The differences can be treated as 'residual anomalies' and can be ascribed to variations in the thickness of the crust.

Figure 3a shows the observed free-air anomalies; Figure 3b shows the topography. The first step is to compute the attraction due to the water layer *abcd*. To do this, one has to specify the coordinates of points which, when joined together consecutively, will define the boundary of this

layer. These are listed in Table 2. It should be noticed that in this particular instance the coordinates of most of the points are determined by the actual corrected soundings and the positions at which these soundings were made.

These data are punched on cards and fed into the IBM 650. It remains to assign a density to this layer. It is found convenient to subtract the densities of all the layers from a constant density of 2.84 gm/cc. (This reduces the density of the crustal layer to zero, and one does not, then, have to make any calculations for it.) Assuming an actual density of 1.03 gm/cc for sea water, the reduced density assigned to the layer *abcd* is  $2.84 - 1.03 = 1.81$  gm/cc.

This value and the coordinates of the points at which the attraction is to be computed are also punched on cards. These latter are points taken at small intervals along the length of the profile. The computed curve is plotted in Figure 3c. This is the correction curve for the water layer. The residual anomalies plotted in Figure 3d are obtained by adding the correction curve to the observed free-air anomalies. Figure 3d may now be used to obtain the variations in crustal thickness.

To the north of the section it is assumed that the water depth is constant at 3.8 km and that the free-air anomalies are constant at -20 mgal. This would require a depth to the Mohorovicic discontinuity of 17.0 km, to be in accord with the standard section of *Worzel and Shurbet* [1955]. Using this as a fixed level, an approximate estimate of thickness is made along the entire section by the  $\sin x/x$  method developed by *Tsuboi and Fuchida* [1938] and *Tomoda and Aki* [1955]. (In the  $\sin x/x$  method the gravity anomalies chosen along a profile at constant intervals are directly attributed to a mass distribution at a fixed depth. The mass distribution is interpreted as being due to the undulations of a surface, which in this case is the Mohorovicic discontinuity). The IBM 650 is used for this calculation also. The estimated depths to the Mohorovicic discontinuity as determined by *Tsuboi's* method are shown as the dashed curve in Figure 3e.

The gravity effect of this estimated crustal thickness was evaluated by the polygon method and compared with the residual anomalies. It was found that there was some disagreement. The depths along the discontinuity were then modified

TABLE 2—List of coordinates used in the deduction of the gravitational effect of the water layer

Distance		Depth	
	km		km
$x_1$	$-\infty$	$z_1$	0.00
$x_2$	$+\infty$	$z_2$	0.00
$x_3$	$+\infty$	$z_3$	4.40
$x_4$	330	$z_4$	4.40
$x_5$	274	$z_5$	4.48
$x_6$	219	$z_6$	4.74
$x_7$	202	$z_7$	4.90
$x_8$	198	$z_8$	4.64
$x_9$	189	$z_9$	3.64
$x_{10}$	184	$z_{10}$	3.10
$x_{11}$	174	$z_{11}$	3.32
$x_{12}$	145	$z_{12}$	3.77
$x_{13}$	102	$z_{13}$	3.75
$x_{14}$	54	$z_{14}$	3.86
$x_{15}$	0	$z_{15}$	3.80
$x_{16}$	$-\infty$	$z_{16}$	3.80
$x_{17}$	$-\infty$	$z_{17}$	0.00

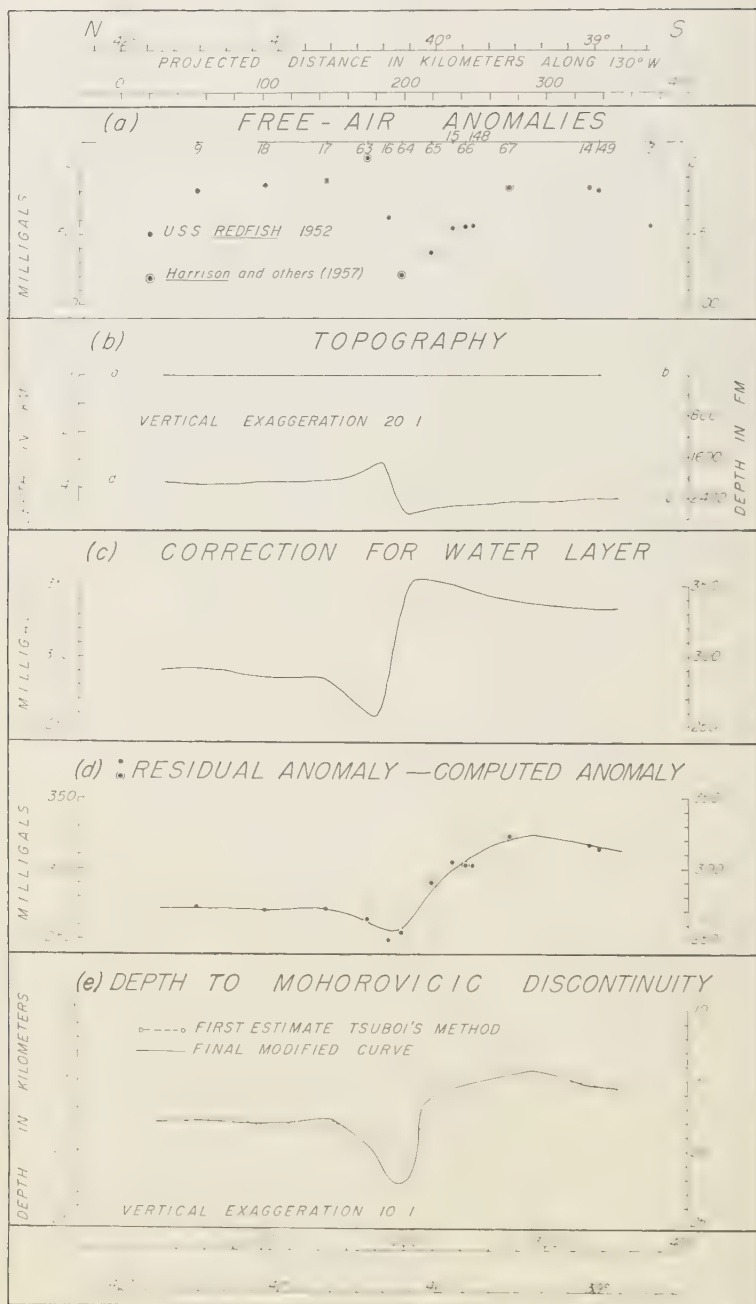


FIG. 3. Various steps involved in the deduction of the depth to the Mohorovicic discontinuity from the free-air anomalies

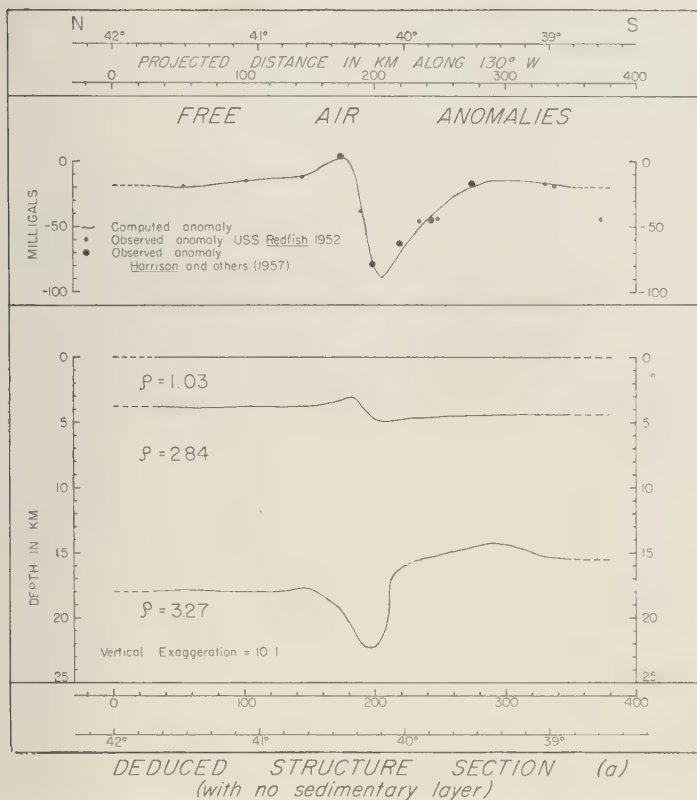


Fig. 4—Free-air anomalies and the deduced crustal structure (a) under the Mendocino fracture zone (with no sedimentary layer)

n such a way as to reduce the disagreement. Alterations of crustal thickness were continued until the computed curve fitted the residual anomalies as well as was desired. The final modified crustal thickness is shown as the solid curve in Figure 3e. The computed residual anomaly curve corresponding to this is shown in Figure 3d and is seen to fit most of the points fairly well. (No attempt was made to fit the gravity value at *Redfish* Station 13, because the unusually large negative anomaly here seems to be of local origin.) It is emphasized here that while Tsuboi's method is very useful in making an approximate estimate of the crustal thickness, it is by no means essential to the solution of the problem. It is quite feasible to guess at an approximate crustal thickness and proceed from there.

Figure 4 shows the final deduced structure (a) and the computed free-air anomaly curve. The

latter is obtained by subtracting the water layer correction curve from the final curve computed to fit the residual anomaly.

The Harrison Stations 61 and 62 and the *Redfish* Stations 146 and 147, which were not used in the profile, can be considered next. The sounding for the Harrison Station 61 is 3330 m, and the free-air anomaly is +13 mgal. The position of this station along the projected profile would be at 137 km south of 42°N. Here the depth read off from the topographic profile is 3750 m, which gives a difference of 420 m. In order to make a rough allowance for this depth difference, we can imagine a uniform plate of water 420 m thick and of infinite areal extent. The gravity effect of this layer (computed in the same way as the simple Bouguer correction), using a density defect of 1.81 gm/cc, is 32 mgal. This has to be subtracted from the value of the



correction due to the water layer at 137 km. The latter can be read in Figure 3c as 286 mgal. Then the approximate residual anomaly at the Harrison Station 61 is  $13 + 286 - 32 = 267$  mgal. (The same number can be obtained from the value given for the Bouguer anomaly by Harrison and others [1957] at Station 61 after allowing for the fact that they have used a different crustal density in making their computations.) This number differs by only  $-5$  mgal from the value of the computed curve at 137 km. Similarly, the differences for the Harrison Station 62 and the Redfish Stations 146 and 147 are  $+1$ ,  $-13$ , and  $-8$  mgal respectively. The comparatively large difference for Station 146 can be attributed to an actual difference in structure, which is expected since the station is at such a great distance from the main Redfish line. This also points to the infeasibility of using both the Redfish and the Harrison stations further north in a single composite profile.

If additional data were available about further layers (as for example a sedimentary layer), the gravity effect of such layers could be 'removed' in the same way as the water layer was removed in the above example. It should be noted that the presence of such layers could alter the deduced section considerably. For instance, a 1-km thickness of sediment of density 2.1 gm/cc would require the crust to be thinner by about 1.7 km in order to reconcile the computed gravity curve with the observed points.

While, in the absence of any seismic determinations of depth and velocity in this region, it was not possible to introduce any complex modifications to the deduced structure, it was felt that an assumption of a sedimentary layer would make the section geologically more probable. Accordingly, the computations were re-done after a sedimentary layer of density 2.1 gm/cc was added to the section at the base of the water layer. This was an average value of density, chosen on the basis of curves given by Nafe and Drake [1957] which relate the compressional velocity and density of ocean sediments. (A value of 2.15 km/sec for the compressional velocity was assumed after Raitt [1956].) The thickness of this sedimentary layer was assumed to be  $\frac{1}{2}$  km over areas of small relief, somewhat greater than  $\frac{1}{2}$  km at the foot of the scarp, and negligible on the scarp itself. The depths to the Mohorovicic discontinuity were obtained in the same way as

in the previous case, the only difference being that the sedimentary layer as well as the water layer had to be removed. Figure 5 shows the deduced crustal section (b) with the assumed sedimentary layer taken into account.

It can be seen that in this case, as in the case without the sedimentary layer, the crust is thicker to the north of the scarp than to the south of it by about three km. The absolute thickness of the crust may be subject to some uncertainty because the values chosen for the densities and the layer thicknesses are based on a standard oceanic section rather than on any seismic determinations in the area. It is of some interest, therefore, to compare the crustal thickness obtained here with those obtained at the closest seismic refraction stations, even though these are at a considerable distance from the gravity profile. The only published determinations available to the authors were Raitt's Stations  $M_1$  and  $M_2$  [Raitt, 1956]. The free-air anomalies in the vicinity of both these stations are of the order of  $-20$  mgal (unpublished Lamont Observatory data). Thus a direct comparison with either end of the gravity profile is in order. At Station  $M_1$  ( $27^\circ 24' N$ ,  $121^\circ 35' W$ ), which lies in Menard and Dietz's Baja California sea mount province, Raitt finds the following structure beneath the water layer: A first layer of thickness 0.26 km and of assumed seismic velocity 2.15 km/sec, a second layer of thickness 0.93 km and velocity 5.88 km/sec, and a third layer of thickness 6.24 km and velocity 6.96 km/sec. Below the Mohorovicic discontinuity the velocity is 8.41 km/sec. Assuming densities of 2.1, 2.6, 2.84 gm/cc for the three layers respectively, and of 3.27 gm/cc for the material below the Mohorovicic discontinuity, and noting that the water depth here is 4.18 km, one can consider the isostatic balance of this section against the section at the south end of the gravity profile. There the water depth is 4.40 km, and a simple calculation shows that a balance with the seismically determined section at  $M_1$  requires a depth to the Mohorovicic discontinuity of 11.65 km, considering a single crustal layer of density 2.84 gm/cc. If allowance is made for an additional sedimentary layer of density 2.1 gm/cc and thickness  $\frac{1}{2}$  km, the depth to the discontinuity is reduced to 10.8 km. A discrepancy is apparent between these crustal thicknesses and those deduced in the gravity sections (a) and (b), the

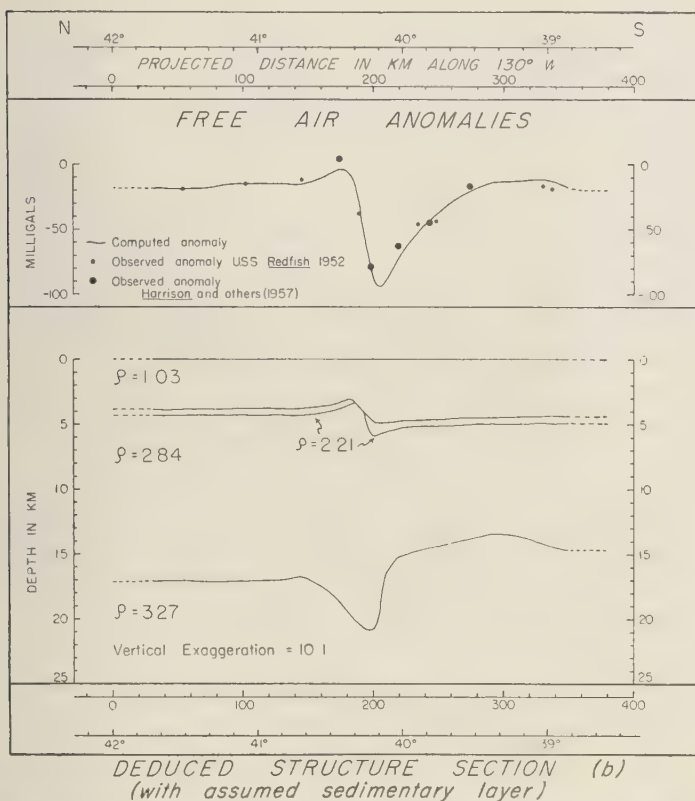


FIG. 5—Free-air anomalies and the deduced crustal structure (b) under the Mendocino fracture zone (with assumed sedimentary layer)

latter giving depths of 15.5 km and 14.65 km respectively. This discrepancy has also been pointed out by Menard [1955]. (His choice of densities and the different water depths in the areas for which he has made the comparison leads him to an even greater discrepancy.) However, if the comparison is made with Raitt's Station  $M_2$  rather than  $M_1$ , the situation is somewhat improved. Proceeding as before, it can be shown that the section at  $M_2$  is 'equivalent' to a section at the south end of the gravity profile having a total crustal thickness of 13.6 km in the absence of a sedimentary layer, and of 12.8 km in the presence of one.

The south end of the gravity profile lies in the abyssal plain. (Menard terms this the "Deep Plains Province.") Since the crustal thickness as deduced here would be in even better agreement with Raitt's seismic determinations at the deep

ocean stations, this might suggest that it is not the deep structure under the abyssal plain that is unusual, as has been proposed by Menard, but that it is the deep structure under the Baja California sea mount province that might be considered so, especially as one approaches the continental edge. More seismic work in both areas is needed before this point can be definitely settled.

The crustal thickness determination under the Mendocino scarp itself is also of interest. We note that, as determined from the gravity computations, the Mohorovicic discontinuity here seems to have a very steep slope just south of the scarp. Some of this may possibly be a spurious effect arising out of the attempt to explain a gravity anomaly of shallow origin by deeper structure. If a volcanic layer of density 2.60 gm/cc and a prismatic shape should be

interposed in the section (with the narrow edge of the prism outcropping at the scarp and the base about 100 km wide extending down to the Mohorovicic discontinuity), this would obviate the necessity of thickening the crust under the escarpment, and would reduce the southern slope of the discontinuity. In any event, it is clear that a mass deficiency exists under the Mendocino escarpment, and it seems probable that the depth at which this deficiency exists is not greater than the depth to the Mohorovicic discontinuity. This should be taken into consideration in any theories explaining the origin of the Mendocino escarpment. Also, it should be noted that the mass deficiency or 'root' is somewhat displaced from the theoretical 'root' which can be computed from the topography on the Airy hypothesis. Undoubtedly, this is the cause of the large isostatic anomalies computed for Stations 64 and 65 by Harrison and others [1957].

So far we have assumed that the anomalies across the Mendocino escarpment have their origin in the crust. It would be of some interest to see if variations of density in the mantle could account for these anomalies equally well.

At the two ends of the profile there is a difference of 44 mgal in the residual anomaly (see

Fig. 3). Let us assume that a constant change of density in the mantle occurs across a vertical interface directly below the Mendocino scarp and extends from a depth of 20 km to 200 km. A simple calculation can then be made to show that a density difference of 0.00583 gm/cc is required to account for the difference of residual anomaly at large distances on either side. Using this density contrast, the gravity curve was computed for the mantle configuration outlined above (Fig. 6). We notice that it fails to fit the anomalies by large amounts in the vicinity of the Mendocino scarp. This further shows that these anomalies must have a shallow origin. While complex modification of the crust, in addition to the density change in the mantle, can probably be made to fit the residual anomaly curve, it is clear that a change of density in the mantle alone cannot account for the observed anomalies.

The actual running time for the computations made in this problem on the IBM 650 was three hours. On a faster computer, the IBM 704, for instance, the running time would be reduced by a factor of at least 30. Of course, a more complicated problem, one involving more layers, for instance, would require correspondingly more time.

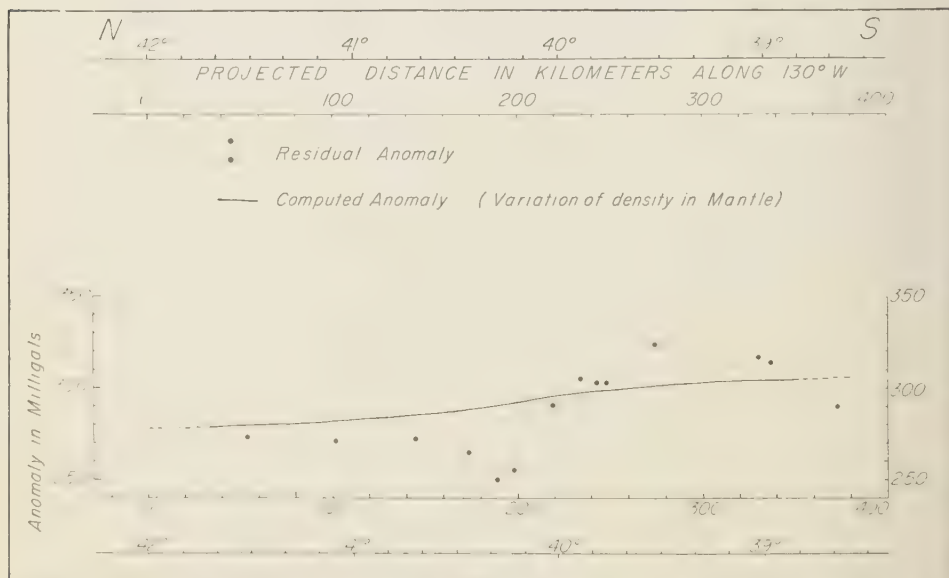


FIG. 6. Attempt to fit the residual anomalies across the Mendocino scarp by a change of density in the mantle (extending to a depth of 200 km)

In conclusion, one must point out that a limitation of this method is the assumed two-dimensionality. Though this is not a serious problem for a large number of structures such as continental margins, ocean trenches, or mountain ranges, it might be a handicap for small scale work. A similar method for rapid calculation of the gravitational attraction of three-dimensional bodies is being developed.

*Acknowledgments*—The authors wish to thank Charles L. Drake, Maurice Ewing, J. E. Nafe, Jack Oliver, and George Sutton for critically reading the manuscript and making many valuable suggestions. Bruce Heezen interested himself keenly in the paper and the authors are indebted to him for many profitable discussions. The computations were made on an IBM 650, made available by the Watson Scientific Computing Laboratory. For that, as well as for the use of other facilities, the authors wish to express their grateful appreciation. Elizabeth S. Skinner helped in making the computations, and Betty Quest, and Annette E. Trefzer drafted the illustrations. The research was supported by the Office of Naval Research, Department of the Navy, under Contract N6-Onr-271, Task Order 8.

#### REFERENCES

- HARRISON, J. C., G. L. BROWN, AND F. N. SPIESS, Gravity measurements in the northeastern Pacific Ocean, *Trans. Amer. Geophys. Union*, **38**, 835-840, 1957.
- HUBBERT, M. KING, A line-integral method of computing the gravimetric effects of two-dimensional masses, *Geophysics*, **13**, 215-225, 1948.
- MENARD, H. W., Deformation of the northeastern Pacific basin and the west coast of North America, *Geol. Soc. Amer. Bul.*, **66**, 1149-1198, 1955.
- MENARD, H. W., AND R. S. DIETZ, Submarine geology of the Gulf of Alaska, *Geol. Soc. Amer. Bul.*, **62**, 1263-1285, 1951.
- MENARD, H. W., AND R. S. DIETZ, Mendocino submarine escarpment, *J. Geol.*, **60**, 266-278, 1952.
- NETTLETON, L. L., *Geophysical prospecting for oil*, McGraw-Hill, 444 pp., 1940.
- NAFE, JOHN E., AND CHARLES L. DRAKE, Variation with depth in shallow and deep water marine sediments of porosity, density and the velocities of compressional and shear waves, *Geophysics*, **22**, 523-552, 1957.
- RAITT, RUSSEL W., Seismic-refraction studies of the Pacific Ocean basin, part I, Crustal thickness of the central equatorial Pacific, *Geol. Soc. Amer. Bul.*, **67**, 1623-1639, 1956.
- SHURBET, G. LYNN, J. LAMAR WORZEL, AND MAURICE EWING, Gravity measurements in the Virgin Islands, *Geol. Soc. Amer. Bul.*, **67**, 1529-1536, 1956.
- TOMODA, YOSHIBUMI, AND KEIITI AKI, Use of the function  $\sin x/x$  in gravity problems, *Japan Acad. Proc.*, **31**, 443-448, 1955.
- TSUBOI, C., AND T. FUCHIDA, Relations between gravity anomalies and the corresponding subterranean mass distribution (II), *Earthq. Res. Inst., Bul.*, **16**, 273-284, 1938.
- VENING MEINESZ, F. A., J. H. F. UMBGROVE, AND PH. H. KUENEN, *Gravity expedition at sea, 1923-1932*, vol. 2, Pub. Netherl. Geod. Comm., 208 pp. 1934.
- WORZEL, J. LAMAR, G. LYNN SHURBET, AND MAURICE EWING, Gravity measurements at sea, 1952 and 1953, *Trans. Amer. Geophys. Union*, **36**, 326-334, 1955.
- WORZEL, J. LAMAR, AND G. LYNN SHURBET, Gravity interpretations from standard oceanic and continental crustal sections, *Geol. Soc. Amer., Spec. Pap.* **62**, 87-100, 1955.

(Communicated manuscript received April 16, 1958, and, as revised, August 15, 1958.)





## Reconciliation of Stokes' Function and Astro-Geodetic Geoid Determinations

W. M. KAULA

*U. S. Army Map Service,  
Washington, D. C.*

**Abstract**—The combined application of astronomic, geodetic, and gravimetric data to computations of the geoid is discussed. The general principle observed is that any adjustment should weight all observations inversely as their variances. Two conditions are imposed: (1) Geoid heights and deflections computed by Stokes' theorem from gravity data must equal those derived by astro-geodetic means. (2) The five harmonics  $P_1$ ,  $P_1^1 \sin \lambda$ ,  $P_1^1 \cos \lambda$ ,  $P_2^1 \sin \lambda$ , and  $P_2^1 \cos \lambda$  must be absent from the adjusted gravity field. The ideal case is discussed, including provision for separate determinations of parameters by independent astronomic methods. Practical modifications are then introduced in turn: treating gravity anomalies as representative of areas; holding geodetic or astronomic observations constant; using a reduced number of astronomic stations; and comparing interpolated points of the astro-geodetic and gravimetric geoids. The most significant discrepancy from the ideal case of most practical solutions made heretofore appears to be in the weighting of the gravimetric data.

**Introduction**—Recently there have been extensive computations of the geoid both by astro-geodetic methods [Lieberman, 1955; Chovitz and Fischer, 1956; Fischer, 1957ab; Bomford, 1957] and by gravimetric methods [Heiskanen, 1957b]. Reconciliation of these determinations is desirable for the purposes of (a) orientation of geodetic systems, (b) determination of ellipsoid parameters, and (c) agreement of gravimetric and astro-geodetic geoid heights and deflections.

Several publications have dealt with the computation of deflections from gravity anomalies and the use of such computations for purpose (a), orientation of geodetic systems. The most extensive study is that of Ledersteger [1951], who interpolated gravimetric deflections from the geoid map of Tanni [1948] and determined the equatorial radius and the orientation of the North American and European geodetic systems from these deflections.

Jeffreys [1948] made more limited use of gravity data by correcting the curvature of the astro-geodetic arcs for harmonics up to the third degree, and a similar application was made by Zhongolovich [1956] of geoid slopes from gravity harmonics, up to the eighth degree, to the determination of ellipsoid dimensions from astro-geodetic data.

The Russian method of gravimetric interpolation between astro-geodetic deflections [Molodenskiy, 1937] accomplishes purpose (c). However,

the method is designed so that purpose (a) will have negligible effect, and assumes that purpose (b) has already been taken care of by other means; therefore an essentially astro-geodetic geoid results. Tengström [1954] also applies gravimetric interpolation between astro-geodetic deflections.

To avoid excessive computational labor on the one hand and systematic error on the other, it is proposed to start with the following general principle, which we shall call the "ideal case" (though even it has some compromises), and then to introduce, one by one, practical modifications which will make the task manageable. The general principle is: Model and datum parameters and observed quantities should be adjusted so that all necessary conditions are met and the weighted sum of the squares of corrections to astronomic, geodetic, and gravimetric observations is a minimum. The weight of an astronomic geodetic, or gravimetric observation should be the inverse of its variance  $\sigma^2$ .

**Preliminary considerations**—The model is an ellipsoid of revolution with (1) parameters considered adjustable in this paper:  $a$ , equatorial radius;  $f$ , flattening; and  $\gamma_e$ , equatorial gravity (alternatively, total mass of the Earth may be considered adjustable); and (2) parameters considered fixed:  $\omega$ , rate of rotation; and  $H$ , precessional constant.

Datums are arbitrary origins or sets of origins

to which observations must be referred, with (1) parameters considered adjustable:  $\phi_0$ ,  $\lambda_0$ , and  $N_0$ , which define the location of a system of geodetic observations with reference to the ellipsoid on which it is computed; and (2) parameters considered fixed:  $\Delta g_0$ , the gravity base station to which a system of gravimetric observations is referred. All major base gravity stations have been tied to the Potsdam system within  $\pm 1$  mgal [Morelli, 1957].

The Laplace azimuths in a geodetic system comprise a set of origins, or reference directions, which fix the orientation  $\alpha_0$  of the geodetic system.

The star catalogue gives stellar mean places which comprise a set of origins, or reference directions, to which astronomic observations are referred.

$x_p$ ,  $y_p$ : the position of the polar axis with respect to the Earth's crust at the time of an astronomic position observation.

Observations are of three types:

(1) Astronomic, in which  $\phi$  and  $\lambda$  define the direction of the vertical at a given point with reference to the fixed stars and are not dependent on any of the parameters considered adjustable.

(2) Geodetic, in which  $\Delta\phi$ ,  $\Delta\lambda$ , and  $\Delta h$ , the distances, azimuths, and differences of elevation between points, can be determined by triangulation, traverse, base line, Laplace azimuth, and leveling operations. When these observations are referred to a datum and ellipsoid, the results are expressed as the differences of position  $\Delta\phi$ ,  $\Delta\lambda$ , and  $\Delta h$ . Hence these values will depend on the parameters considered adjustable,  $\alpha$ ,  $f$ ,  $\phi_0$ ,  $\lambda_0$ , and  $N_0$ .

(3) Gravimetric, in which  $\Delta g$ , the intensity of gravity is expressed as a departure from the model,  $g - \gamma$ , and hence is dependent on the parameters  $\gamma_e$  and  $f$ . It is usually measured relatively, as a difference in gravity between points; hence  $\Delta g$  is also dependent on the parameter  $\Delta g_0$ . Gravity can only be observed at a finite number of points; however, to use  $\Delta g$ , it must be integrated over a surface.  $\Delta g$  must therefore be represented either by mean values applying to arbitrary areas (for example, squares of equal latitude and longitude in extent), or by an arbitrary number of polynomials (for example, spherical harmonics). Either method is feasible for large geodetic systems, but this discussion shall be confined to methods using Stokes'

function, which can be applied to small systems as well. Hence,  $\Delta g$  shall be taken herein to be the mean anomaly for an arbitrary area. The uncertainty of  $\Delta g$ , expressed by standard error  $\sigma$  or variance  $\sigma^2$ , is thus mainly an error of representation, which measures the failure of point anomalies to represent an area. To minimize complications, it is assumed herein that the anomalies are free-air, with topographic or elevation corrections where appropriate.

Conditions—Two conditions must be met.

Condition I. If the results of the observations are to refer to the same datum and model, the following two forms of the condition must be met.

Form A. The difference in position between any two points obtained by geodetic observations must equal the difference between their astronomic positions minus their corresponding gravimetrically computed deflections.

Geod.	Astr.	Grav.	Astr.	Grav.
-------	-------	-------	-------	-------

$$\Delta\phi_{ab} = (\phi_a - \xi_a) - (\phi_b - \xi_b)$$

$$\Delta\lambda_{ab} = (\lambda_a - \eta_a \sec \phi_a) - (\lambda_b - \eta_b \sec \phi_b),$$

or

$$\left. \begin{aligned} \Delta\phi_{ab} &= \left( \phi_a - \sum_i c_{ai}' \cos \alpha_{ai} \Delta g_i \right) \\ &\quad - \left( \phi_b - \sum_i c_{bi}' \cos \alpha_{bi} \Delta g_i \right) \\ \Delta\lambda_{ab} &= \left( \lambda_a - \sec \phi_a \sum_i c_{ai}' \sin \alpha_{ai} \Delta g_i \right) \\ &\quad - \left( \lambda_b - \sum_i c_{bi}' \sin \alpha_{bi} \Delta g_i \right), \end{aligned} \right\} \quad (1)$$

where

$$c_{ai}' = \frac{\csc 1''}{2\pi G} \int_{a_i} \frac{df}{d\psi_a} dq$$

$dq$  is an element of area, and  $df/d\psi$  is the Vening Meinesz function; or, with sufficient approximation in many cases,

$$c_{ai}' = \frac{\csc 1''}{2\pi G} \left( \frac{df}{d\psi} \right)_{a_i} A_i,$$

where  $A_i$  is the area, on the unit sphere, represented by  $\Delta g_i$ .

Form B. The increment in elevation of the geoid above the reference model between two points, obtained by integrating astro-geodetic deflections over a line between the two points, must equal the difference of the gravimetrically computed geoid elevations.

$$\begin{aligned}
 N_a - N_b &= \int_b^a [(\phi_a - \phi_g) \cos \alpha \\
 &\quad + (\lambda_a - \lambda_g) \cos \phi \sin \alpha] dl \\
 &= \sum_i c_{ai} \Delta g_i - \sum_i c_{bi} \Delta g_i \quad (2)
 \end{aligned}$$

where  $c_{ai} = (R/2\pi G) \int_{a_i} f(\psi)_a dq$ , the integral of Stokes' function over the area represented by  $\Delta g_i$ ; or, with sufficient approximation in many cases,  $c_{ai} = (R/2\pi G) f(\psi)_{a_i} A_i$ .

Forms A and B of Condition I are not independent because a line integral of an analytic function taken between two points on a continuous surface is the same, regardless of the route taken between the two points. In this case, integrating the deflection over a line following a chain of geodetic observations between two points should give the same answer as splitting up the deflection into  $n$  parts and integrating them over  $n$  lines, each line of two segments  $\psi_{ai}$ ,  $\alpha_{ai}$  and  $\psi_{bi}$ ,  $\alpha_{bi}$ . For theoretical development, Form A of Condition I is the more convenient, since it makes the most direct comparison of observations. For actual application, Form B might be superior, since the integration of the astro-geodetic deflections, as well as the gravity data, would impose closure conditions and smooth out the effect of any erroneous or non-typical deflections. The more easily grasped Form A will be used in studying the ideal case, but Form B will be discussed as a practical modification.

Condition II. If the center of gravity and inertial axes of the geoid and of the reference model are to coincide, then  $\Delta g$  cannot contain the three first-degree harmonics, or two of the second-degree harmonics:

$$\sin \phi \cos \phi \begin{cases} \sin \lambda \\ \cos \lambda \end{cases}$$

[Lambert, 1957]. If we sum gravity, expressed by mean anomalies representative of areas, over any part of the sphere, the result must be equal to gravity expressed as a harmonic summed over the same part of the sphere: partial

$$\begin{aligned}
 &\sum_i \Delta g_i A_i \\
 &= \int_s \sum_{n=0}^{\infty} \sum_{m=0}^n P_n^m(\sin \phi) \begin{cases} a_{mn} \sin m\lambda \\ b_{mn} \cos m\lambda \end{cases} d\sigma \quad (3)
 \end{aligned}$$

Now if

$$P_s^r(\sin \phi) \begin{cases} \sin r\lambda \\ \text{or} \\ \cos r\lambda \end{cases}$$

is a harmonic which cannot exist in the Earth's gravitational field, then

$$\begin{aligned}
 &\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} P_s^r(\sin \phi) \begin{cases} \sin r\lambda \\ \text{or} \\ \cos r\lambda \end{cases} \sum_{n=0}^{\infty} \sum_{m=0}^n P_n^m(\sin \phi) \\
 &\quad \cdot \begin{cases} a_{mn} \sin m\lambda \\ b_{mn} \cos m\lambda \end{cases} \cos \phi d\phi d\lambda = 0, \quad (4)
 \end{aligned}$$

since

$$\int P_s^r \begin{cases} \cos r\lambda \\ \text{or} \\ \sin r\lambda \end{cases} P_n^m \begin{cases} \cos m\lambda \\ \text{or} \\ \sin m\lambda \end{cases} d\sigma = 0,$$

if  $m \neq r$  or  $n \neq s$ ; and  $a_{mn} = 0$ ,  $b_{mn} = 0$  if  $r = m$  and  $s = n$ . So then also:

$$\sum_i \Delta g_i A_i P_s^r(\sin \phi_i) \begin{cases} \cos r\lambda_i \\ \text{or} \\ \sin r\lambda_i \end{cases} = 0. \quad (5)$$

For the five inadmissible harmonics of the gravity field this simplifies to

$$\sum_i \Delta g_i A_i \cos \phi_i = 0,$$

$$\sum_i \Delta g_i A_i \sin \phi_i \sin \lambda_i = 0,$$

$$\sum_i \Delta g_i A_i \sin \phi_i \cos \lambda_i = 0, \quad (6)$$

$$\sum_i \Delta g_i A_i \sin \phi_i \cos \phi_i \sin \lambda_i = 0,$$

$$\sum_i \Delta g_i A_i \sin \phi_i \cos \phi_i \cos \lambda_i = 0.$$

*Adjustment of the ideal case*—We introduce corrections to the observations in (1),

$$\begin{aligned}
 &\phi_a + \delta\phi_a - \sum_i c_{ai}' \cos \alpha_{ai} (\Delta g_i + \delta\Delta g_i) \\
 &- (\phi_b + \delta\phi_b) + \sum_i c_{bi}' \cos \alpha_{bi} (\Delta g_i + \delta\Delta g_i) \\
 &= \Delta\phi_{ab} + \delta\Delta\phi_{ab} \\
 &+ \frac{\partial \Delta\phi}{\partial h} [N + \delta N + \frac{1}{2}(\delta h_a + \delta h_b)], \quad (7)
 \end{aligned}$$



and a similar equation for the difference in longitude. Included here is the assumption that the geoid elevation  $N$  was not allowed for in computing the triangulation. The correction to elevation enters Condition IA to a very minor degree, being multiplied by the small coefficient  $\partial\Delta\phi_{ab}/\partial h$ . Since level observations are very precise as well, further mention of  $\delta h$ ,  $\Delta h$ , or  $\delta\Delta h$  will be omitted. The  $N$  and  $\delta N$  cannot be neglected, and the condition equation will be modified by the assumption that they will be the mean of the gravimetrically computed  $N$  and  $\delta N$  at the end points.

$$\begin{aligned} & (\phi_a + \delta\phi_a) - (\phi_b + \delta\phi_b) \\ & + \sum_i (c_{bi}' \cos \alpha_{bi} - c_{ai}' \cos \alpha_{ai})(\Delta g_i + \delta\Delta g_i) \\ & = \Delta\phi_{ab} + \delta\Delta\phi_{ab} \\ & + \frac{\partial\Delta\phi_{ab}}{\partial h} \left[ \frac{1}{2} \sum_i (c_{ai} + c_{bi})(\Delta g_i + \delta\Delta g_i) \right]. \quad (8) \end{aligned}$$

We introduce corrections to parameters,

$$\begin{aligned} & (\phi_a + \delta\phi_a) - (\phi_b + \delta\phi_b) \\ & + \sum_i (c_{bi}' \cos \alpha_{bi} - c_{ai}' \cos \alpha_{ai}) \left( \Delta g_i + \delta\Delta g_i \right. \\ & \left. + \frac{\partial\Delta g_i}{\partial f} \delta f + \frac{\partial\Delta g_i}{\partial \gamma_e} \delta \gamma_e \right) \\ & = \Delta\phi_{ab} + \delta\Delta\phi_{ab} + \frac{\partial\Delta\phi_{ab}}{\partial \phi_0} \delta\phi_0 \\ & + \frac{\partial\Delta\phi_{ab}}{\partial \lambda_0} \delta\lambda_0 + \frac{\partial\Delta\phi_{ab}}{\partial N_0} \delta N_0 + \frac{\partial\Delta\phi_{ab}}{\partial f} \delta f \\ & + \frac{\partial\Delta\phi_{ab}}{\partial h} \left[ \delta a + \frac{1}{2} \sum_i (c_{ai} + c_{bi}) \left( \Delta g_i + \delta\Delta g_i \right. \right. \\ & \left. \left. + \frac{\partial\Delta g_i}{\partial f} \delta f + \frac{\partial\Delta g_i}{\partial \gamma_e} \delta \gamma_e \right) \right], \quad (9) \end{aligned}$$

and a similar equation for the difference in longitude. We rearrange, with all the unknowns on one side and the terms which can be computed beforehand on the other,

$$\begin{aligned} & \delta\phi_a - \delta\phi_b - \delta\Delta\phi_{ab} + \left[ \sum_i (c_{bi}' \cos \alpha_{bi} \right. \\ & \left. - c_{ai}' \cos \alpha_{ai} - \frac{1}{2} \frac{\partial\Delta\phi_{ab}}{\partial h} (c_{ai} + c_{bi}) \right] \end{aligned}$$

$$\begin{aligned} & \times \left[ \Delta g_i + \frac{\partial\Delta g_i}{\partial f} \delta f + \frac{\partial\Delta g_i}{\partial \gamma_e} \delta \gamma_e \right] \\ & - \frac{\partial\Delta\phi_{ab}}{\partial \phi_0} \delta\phi_0 - \frac{\partial\Delta\phi_{ab}}{\partial \lambda_0} \delta\lambda_0 - \frac{\partial\Delta\phi_{ab}}{\partial N_0} \delta N_0 \\ & - \frac{\partial\Delta\phi_{ab}}{\partial h} \delta a - \frac{\partial\Delta\phi_{ab}}{\partial f} \delta f = k_{ab}, \quad (10) \end{aligned}$$

where  $k_{ab}$  is the sum of the known terms. In this condition equation, we have assumed that a single chain or block of geodetic observations connects the two astronomic stations  $a$  and  $b$ . Actually, there may be more than one geodetic chain included in the condition equation. For example, if three astronomic stations are connected by geodetic chains in the form of a "T", there will be two condition equations; the stem of the "T" will be included in both equations and the arms each in one. In an entire geodetic system, the number of independent condition equations will be the minimum number of equations necessary to include each chain or block of geodetic observations at least once.

Condition II equations for the inadmissible harmonics will be only five in number and will include only the corrections  $\delta\Delta g$ ,  $\delta f$ , and  $\delta\gamma_e$ .

Ideally, we may solve for all of the corrections by one large least-squares adjustment. A repetition of some of the mathematical steps may assist in understanding subsequent modifications of geodetic significance. The treatment is essentially that of *Arley and Buch* [1950, pp. 196-198].

We assume a total of  $m$  corrections to observations,  $r$  parameters, and a total of  $n$  condition equations, including both Condition I and Condition II. Column matrices are formed,

$$X_{m \times 1} = \begin{Bmatrix} \delta\phi's \\ \delta\lambda's \\ \delta\Delta\phi's \\ \delta\Delta g's \end{Bmatrix}, \quad Y_{r \times 1} = \begin{Bmatrix} \delta\phi_0's \\ \delta\lambda_0's \\ \delta N_0's \\ \delta a \\ \delta f \\ \delta\gamma_e \end{Bmatrix},$$

and

$$K_{n \times 1} = \{k_{ab}'s\}, \quad (11)$$

where the  $k_{ab}$  are the constants, or known terms, in the condition equations. The conditions are expressed by a matrix equation,

$$K = BX + CY, \quad (12)$$

where the elements of  $B$  will be the various coefficients of the corrections to observations in the condition equations:  $+1$  or  $-1$  or (usually)  $0$  for  $\delta\phi$ ,  $\delta\lambda$ ,  $\delta\Delta\phi$ , and  $\delta\Delta\lambda$ ;

$$\sum \left\{ c_{bi}' \cos \alpha_{bi} - c_{ai}' \cos \alpha_{ai} - \frac{1}{2} \frac{\partial \Delta\phi_{ab}}{\partial h} (c_{ai} + c_{bi}) \right\}$$

or  $A_i \sin \phi_i \sin \lambda_i$ , etc. for  $\delta\Delta g$ . The elements of  $C$  will be the coefficients of the corrections to parameters: that is, partial derivatives of the geodetic and gravimetric observations with respect to the parameters.

We use  $r$  of (12) to solve for the parameters  $Y$  in terms of the observations  $X$ . This leaves  $n - r$  independent conditions:

$$K' = B' X \quad (13)$$

The primes hereafter will indicate a condition matrix remaining after the parameters have been eliminated. The sum of the squares of the corrections to be minimized is

$$S = X^* P X, \quad (14)$$

where  $X^*$  is the transpose of  $X$ , and  $P$  the weight matrix. The customary method of correlates yields

$$X = P^{-1} B' (B' P^{-1} B')^{-1} K', \quad (15)$$

which gives us the corrections to the observations.  $Y$ , the corrections to the parameters, can be obtained by using the same  $r$  condition equations used to eliminate them in the first step.

Now note that in the condition equation (7) we can express

$$\begin{aligned} \delta \xi_a &= \sum_i c_{ai}' \cos \alpha_{ai} \delta \Delta g_i \\ \delta N_a &= \sum_i c_{ai} \delta \Delta g_i \end{aligned} \quad (16)$$

$$\frac{\partial \xi_a}{\partial f} \delta f = \sum_i c_{ai}' \cos \alpha_{ai} \frac{\partial \Delta g_i}{\partial f} \delta f$$

In other words, the  $\xi$ ,  $\eta$ , and  $N$  are all linear transformations of the  $\Delta g$ 's. If a set of linearly transformed unknowns is adjusted to a least-squares condition, then the solution will be fully consistent with a least-squares adjustment of the original unknowns [Tienstra, 1947]. The principal

caution we must observe is that the variances and covariances of the transformed unknowns are based on the variances of the original unknowns. We form a new set of unknowns,

$$Z = \begin{Bmatrix} \delta\phi's \\ \delta\lambda's \\ \delta\Delta\phi's \\ \delta\Delta\lambda's \\ \delta\xi's \\ \delta\eta's \\ \delta N's \end{Bmatrix}, \text{ and, say } Z = D X. \quad (17)$$

The elements of  $D$  will be 1 where the elements of  $Z$  and  $X$  are the same; they will be 0 in many places; and they will be the Stokes' functions  $c_{ai}$  and  $c_{ai}'$  in others. The advantage of using  $Z$  is that we can arrange it so that all the elements, for example,  $\delta\phi$ ,  $\delta\xi$ ,  $\delta N$ , and  $\delta\Delta\phi$ , pertaining to a particular condition are close together. The condition equations (7) are now much easier to form, since the Stokes' functions are no longer required. The five inadmissible harmonics of Condition II cause little more difficulty. The Stokes' integration automatically eliminates the inadmissible first-degree harmonics, with distortion introduced only if the gravity anomalies are improperly weighted [Hirvonen, 1955]; hence we can ignore these three conditions in the first stage. The inadmissible second-degree terms of gravity will give rise to similar terms in the form of the geoid

$$\frac{R}{G(n-1)} P_2^1(\sin \phi) \begin{Bmatrix} a_{12} & \cos \lambda \\ b_{12} & \sin \lambda \end{Bmatrix},$$

hence imposing conditions the same as for  $\Delta g$ ,

$$\sum_i N_i A_i \sin \phi_i \cos \phi_i \begin{Bmatrix} \sin \lambda_i \\ \cos \lambda_i \end{Bmatrix} = 0, \quad (18)$$

where  $A_i$  is the area of representation for  $N_i$ . In the ideal case, we must compute  $N_i$  the world over from  $\Delta g$  and impose this condition in the first stage. In the practical case, where we are comparing the gravimetric and astro-geodetic geoids only where the latter exist, we can safely assume that any inadmissible harmonic introduced would be absorbed by an adjustment to  $N_i$  in one of the vast areas of the ocean devoid of gravity observations.

We can solve for  $Z$  and  $Y$  in the same manner as is described for  $X$  in (15) of the one-stage

solution, with appropriate changes in the coefficient and weight matrices:

$$\begin{aligned} K_{(n-5) \times 1} &= B(D^*D)^{-1}D^*Z + CY \\ &= EZ + CY, \quad (19) \end{aligned}$$

the  $(n - 5)$  indicating that we are not including the inadmissible harmonics conditions. The new square sum is  $Z^*QZ$ ; computation of  $Q$  is discussed below. After solving for  $Z$ , the values of  $\delta\Delta g$  may be computed by a least-squares adjustment in which the values of  $\delta\xi$ ,  $\delta\eta$ ,  $\delta N$ , and the three or five inadmissible harmonics not imposed in the first stage, are the conditions. The new values of  $\Delta g$  will apply to an ellipsoid as adjusted in the first stage.

The variances in the geodetic observations will be due principally to variances in the angles measured in triangulation and, for each block or chain between astronomic points, must be estimated by methods such as those described by Bomford [1952; pp. 123-127]. The astronomic variances will be purely those of the position observations themselves. Neither the geodetic nor the astronomic observations should involve any correlation; therefore the weights will simply be the inverses of the variances.

The variance of  $\Delta g$ , defined here as the mean anomaly for an arbitrary area, is a function of the size of the area and the distribution of observations within it, as described by Hirvonen [1956] and Kaula [1957]. In cases where we have areas empty of observations, we must allow for covariance as well. There will be covariance between empty areas within  $30^\circ$  of each other and also between an empty area with an interpolated anomaly and those areas from whose anomalies the interpolation was made. To determine the weights when the covariance exists, we must form a variance-covariance matrix  $V$  and invert the entire matrix to get a weight matrix  $R = V_G^{-1}$ .

In using the two-stage solution, the variance of the deflections must be based on the variance of the gravity anomalies. We take that submatrix of  $Q$  which pertains to the  $\delta\xi$ ,  $\delta\eta$ ,  $\delta N$  and compute their weights separately;

$$Q_1 = V_D^{-1} = (D_1 V_G D_1^*)^{-1}, \quad (20)$$

where  $D$  is a sub-matrix of  $D$ .  $Q_1$  and  $V_D$  are solid matrices, since  $\xi$ ,  $\eta$ ,  $N$  are correlated as being functions of the same  $\Delta g$ 's.

The weights used herein necessarily have dimensions, since we include in the same square sum  $X^*PX$  the quantities  $\delta\phi$ ,  $\delta\Delta\phi$  in seconds of arc and  $\delta\Delta g$  in milligals.

Heretofore we have assumed parameters to be, in effect, unobserved quantities with infinite variance and zero weight. If we have a measurement of a parameter which is obtained independently of the other elements adjusted, then we should use this measurement and ascribe to it a weight which is the inverse of its variance. Instances which can occur:

The equatorial radius  $a$ . Measurement of the distance to the moon by radar [Yaplee, Bruton, Craig, and Roman, 1958] combined with the dynamical parallax gives a value for the equatorial radius independent of geodetic methods. [O'Keefe and Anderson, 1952].

The flattening  $f$ . The precession of the node of a satellite orbit gives an independent value of  $f$ . The variance will depend in part on the effect of higher degree zonal harmonics. Also if the Earth is assumed to be in hydrostatic equilibrium, then the flattening can be determined from the precessional constant  $H$ , which has been assumed to be fixed in this paper. The variance, and hence the weight, of  $f$  would depend mainly on the uncertainty of the hydrostatic assumption [Jeffreys, 1948].

The datum parameters  $\phi_0$  and  $\lambda_0$ . A connection between two datums could be made by a method with an appreciable uncertainty, such as occultation or satellite observations. This connection,  $\Delta\phi_0$  and  $\Delta\lambda_0$ , may be treated as an observation, with a weight the inverse of its variance, and two additional condition equations,  $\Delta\phi_0 = (\phi_0 - \phi'_0)$  and  $\Delta\lambda_0 = (\lambda_0 - \lambda'_0)$ .

*Practical modifications*—Mention of the size of the areas represented by the gravity 'observations' has been avoided thus far in this paper. It has a lower limit determined by the number of variables which can be conveniently handled in a computation and by the significance of the resulting corrections. If we try to make the areas variable in proportion to their effects on the deflection conditions, we get trivially small areas close to the astronomic stations combined with the wide expanses distant from the stations. The following procedure appears to reconcile the difficulty:

(1) Establish a lower limit on the size of the area used; for example,  $1^\circ \times 1^\circ$  squares.

(2) Estimate the variance of a deflection from a 'local' area out to, say,  $3^\circ$  from the station, using a system of zones and sectors,  $\sigma_T^2$ .

(3) Compute the variance of the deflection due to the variances of the mean anomalies of the  $1^\circ \times 1^\circ$  squares covering the same area,  $\sigma_M^2$ .

(4) The difference of the two,  $\sigma_L^2 = \sigma_T^2 - \sigma_M^2$ , will be due to variations of the gravity field within the  $1^\circ \times 1^\circ$  squares. These variations can be lumped together as one term in the adjustment  $\delta\xi_L$  or  $\delta\eta_L$  for each deflection  $\xi$  or  $\eta$  with a weight  $1/\sigma_L^2$ .

When this method is used, the lower limit on the square size is not dependent on the nearby deflection point at all; it is dependent on the next nearest point and involves considerations taken up in the next paragraph.

The compartments or areas to be used in computing the effect of the known gravity field, needed for the constant terms in the condition equations, should be those used by the Gravity Project of Ohio State University [Uotila, 1957], that is, the system of zones and sectors developed by Rice [1952] out to  $\psi = 3^\circ$ ;  $1^\circ \times 1^\circ$  squares for  $3^\circ < \psi < 20^\circ$ ; and  $5^\circ \times 5^\circ$  squares for  $20^\circ < \psi < 180^\circ$ . This makes maximum utilization of data already collected and analyzed and of established procedures.

For the areas whose mean anomalies must be adjusted, however, a different breakdown may be desirable. The method just described takes care of the area within  $3^\circ$  by substituting  $1^\circ \times 1^\circ$  squares plus the term for 'local effect'. However, for more distant areas we may want to use larger squares in order to reduce the number of unknowns which must be found in the solution. Eq. (15) indicates that the correction taken by each unknown is roughly proportional to variance  $\times$  (effect on conditions)  $\approx$  variance  $\times$  (effect on deflections). This suggests a criterion for selecting the size of square to be used without introducing excessive error.

The size of square, with side length  $L$ , to be represented by a correction to its mean anomaly should be such that using it in place of four squares of side length  $L/2$  will not cause an error greater than a specified amount  $[\epsilon'']^2$  in computing the variance of a deflection, at distance  $\psi_m$ , from the variance of the mean gravity anomalies. To examine this criterion, let us assume that the variances of the two inner squares,  $V_i$ , are equal, and that the variances

of the two outer squares are  $V_0$ .  $V_i$  and  $V_0$  will be functions, respectively, of  $s_i$  and  $s_0$ , the spacings of gravity observations in the inner and outer squares. We set

$$I_i = \int_{\psi_i}^{\psi_0} \frac{df}{d\psi} \sin \psi \, d\psi,$$

$$I_i = \int_{\psi_i}^{\psi_m} \frac{df}{d\psi} \sin \psi \, d\psi,$$

$$I_0 = \int_{\psi_m}^{\psi_0} \frac{df}{d\psi} \sin \psi \, d\psi$$

and make  $V_i$  the variance of the mean anomaly of the large square; then

$$[\epsilon'']^2 = \left[ \frac{\csc 1''}{2\pi G} \right]^2 \{ V_i I_i^2 (\alpha_r - \alpha_i)^2 - V_i I_i^2 (\alpha_r - \alpha_i)^2 - V_0 I_0^2 (\alpha_r - \alpha_i)^2 \}. \quad (21)$$

Now

$$\alpha_r - \alpha_i = \frac{(\psi_0 - \psi_i)}{\psi_m} = \frac{L}{\psi_m},$$

and

$$V_i = \frac{s_i^2}{L \times \frac{L}{2}} F_i^2 = \frac{2s_i^2}{L^2} F_i^2, \quad V_0 = \frac{2s_0^2}{L^2} F_0^2,$$

where  $F$  is the standard mean error of the weighted mean of an observed anomaly and is zero for a square of side length  $s$  [Hirvonen, 1956]. If  $V_i = V_i V_0 / (V_i + V_0)$  is used, there results

$$[\epsilon'']^2 = \left[ \frac{\csc 1''}{2\pi G} \right]^2 \left[ \frac{L}{\psi_m} \right]^2 \left\{ \frac{4s_i^2 s_0^2}{L^4} F_i^2 F_0^2 - \frac{2s_i^2}{L^2} F_i^2 + \frac{2s_0^2}{L^2} F_0^2 \right. \\ \left. - I_i^2 - \frac{2s_i^2}{L^2} F_i^2 I_i^2 - \frac{2s_0^2}{L^2} F_0^2 I_0^2 \right\} \\ = \left[ \frac{\csc 1''}{2\pi G} \right]^2 \left[ \frac{1}{\psi_m} \right]^2 \left\{ \frac{2s_i^2 s_0^2 F_i^2 F_0^2}{s_i^2 F_i^2 + s_0^2 F_0^2} - I_i^2 - 2s_i^2 F_i^2 I_i^2 - 2s_0^2 F_0^2 I_0^2 \right\}. \quad (22)$$

The side length  $L$  thus enters only through the integrals  $I_i$ ,  $I_i$ , and  $I_0$  and would have to be



found by trial and error. The equation also emphasizes that the greatest error will occur when  $V_i$  and  $V_0$ , or rather  $s_i$  and  $s_0$ , differ appreciably. For practical cases, the following rules are suggested:

(1) After the astronomic points have been selected, make the smallest size square for gravity anomalies the largest such that the local effects,  $\delta\xi_L$  and  $\delta\eta_L$ , of no two astronomic stations overlap to any appreciable extent.

(2) Compute the  $[\epsilon'']^2$  for some typical cases of the squares so selected.

(3) Select the remaining areas so that the  $[\epsilon'']^2$  are of the same magnitude and so that the total number gives a reasonable number of degrees of freedom for adjustment; say, two or three times as many gravity anomaly areas as there are astronomic stations.

In the discussion of the two-stage solution, we included  $\delta N$  as an element to be adjusted in the first stage. However,  $\delta N$  is only weakly connected to the other elements in the condition equations, since it is multiplied by the small factor  $\partial\Delta\phi/\partial h$  or  $\partial\Delta\lambda/\partial h$ ; and  $\delta N$  in turn is a much weaker condition on the values of  $\delta\Delta g$  in the second stage than are the  $\delta\xi$  and  $\delta\eta$ . It probably would do very little damage to: (1) assume the  $\delta N = 0$  in the first stage, (2) solve for  $\delta\Delta g$  to fit the  $\delta\xi$  and  $\delta\eta$  in the second stage, (3) compute the  $\delta N$  resulting from the  $\delta\Delta g$ , (4) apply these  $\delta N$  values to the constants in the first stage conditions, and (5) repeat the first and second stages, finding  $\delta'\xi$ ,  $\delta'\eta$ , and  $\delta'\Delta g$ , but setting  $\delta'N = 0$ .

The geodetic connections  $\Delta\phi$  and  $\Delta\lambda$  will often have appreciably smaller variances than the astronomic observations, particularly if the connections are less than 500 km long. In even more cases, the  $\delta\Delta g$  will take the bulk of the adjustment. However, for distances on the order of 4000 km or more we would not get the full benefit of the adjustment unless  $\delta\Delta\phi$  and  $\delta\Delta\lambda$  were included, because in this case appreciable changes in position might show up at the ends of long arcs such as that down the west side of South America.

A benefit of holding the lengths and azimuths of the geodetic connections constant would be a reduction in the number of condition equations. It is no longer necessary to include each geodetic link at least once; instead, the number of independent conditions will be the minimum number of links needed to connect all of the astronomic

stations to the geodetic system. For example, in a rectangular system with  $m \times n$  astronomic points at the junctions,  $4mn - 2(m + n)$  condition equations are needed if  $\delta\Delta\phi$  and  $\delta\Delta\lambda$  are included; if we hold  $\Delta\phi$  and  $\Delta\lambda$  constant, then  $2mn - 2$  conditions are needed, which amounts to a reduction of almost half.

Following the logic of the two previous paragraphs a step further, the astronomic  $\phi$  and  $\lambda$  can also be fixed, leaving the values of  $\Delta g$  and the parameters as the only elements to be adjusted. This adjustment would be similar to the second stage of the two-stage adjustment. The condition equation for deflection in latitude at point  $a$  is

$$\begin{aligned} \xi_a + \frac{\partial \xi_a}{\partial a} \delta a + \frac{\partial \xi_a}{\partial f} \delta f + \frac{\partial \xi_a}{\partial \phi_0} \delta \phi_0 \\ + \frac{\partial \xi_a}{\partial \lambda_0} \delta \lambda_0 + \frac{\partial \xi_a}{\partial N_0} \delta N_0 = \\ \delta \xi_L + \sum_i c_{ai}' \cos \alpha_{ai} \left( \Delta g_i + \delta \Delta g_i \right. \\ \left. + \frac{\partial \Delta g_i}{\partial \gamma_e} \delta \gamma_e + \frac{\partial \Delta g_i}{\partial f} \delta f \right). \quad (23) \end{aligned}$$

If we collect the knowns on one side and the unknowns on the other

$$\begin{aligned} \sum_i c_{ai}' \cos \alpha_{ai} \delta \Delta g_i \\ + \left( \sum_i c_{ai}' \cos \alpha_{ai} \frac{\partial \Delta g_i}{\partial f} - \frac{\partial \xi_a}{\partial f} \right) \delta f \\ + \sum_i c_{ai}' \cos \alpha_{ai} \frac{\partial \Delta g_i}{\partial \gamma_e} \delta \gamma_e + \delta \xi_L - \frac{\partial \xi_a}{\partial a} \delta a \\ - \frac{\partial \xi_a}{\partial \phi_0} \delta \phi_0 - \frac{\partial \xi_a}{\partial \lambda_0} \delta \lambda_0 - \frac{\partial \xi_a}{\partial N_0} \delta N_0 = k_a, \quad (24) \end{aligned}$$

with a similar equation in longitude, or in  $\eta$ . In this solution, we have lost the  $(\partial\Delta\phi/\partial h)(N + \delta N)$  terms. However, we have lost scale in setting the geodetic observations constant; therefore the only change necessary will be to subtract the mean height of the geoid,  $(N + \delta N)$ , over the area of adjustment from the equatorial radius  $a$  at the end of the computation [Ledersteger, 1951, p. 114]. In addition, there still are the five inadmissible harmonic conditions.

It is not practicable to utilize all of the first-order astronomic positions available because of (1) the labor of computing the deflections gravimetrically,  $\sum_i c_{ai}' \Delta g_i$ , and (2) the necessity of

either using much smaller gravity anomaly areas or contriving some means whereby certain local effect terms,  $\delta\xi_L$  and  $\delta\eta_L$ , can apply to more than one astronomic station without being doubly or overweighed.

If we reduce the number of astronomic stations but retain  $\delta\phi$  and  $\delta\lambda$  as variables to be found in the adjustment, the weight of any astronomic station should be the sum of the weights of the stations it represents. This increase in weight of the astronomic stations when a reduced number is taken, leads to the method described in the preceding paragraph: ascribe infinite weight to the astronomic stations, that is, hold them constant.

Another question is whether to space the selected stations uniformly over the geodetic system or to distribute them in proportion to the number of stations actually observed. When the astronomic observations are taken as constant, the two procedures should give about the same answer, with more accurate detailing of the gravity field from the proportionate distribution where astronomic stations are dense.

The use of representative astronomic points suggests the next step. Interpolated deflections are used at regular intervals, for example, every  $5^\circ$  intersection in latitude and longitude. However, in this case we cannot treat the astrogeodetic deflections as constant. An interpolated deflection will, in general, have an appreciable uncertainty, and will require consideration of the covariance of deflections to evaluate.

Another method is to use the geoidal heights  $N$  at regular intervals (or averaged over areas). The condition equation would be

$$\begin{aligned} N_a + \delta N_a + \frac{\partial N_a}{\partial N_0} \delta N_0 + \frac{\partial N_a}{\partial \phi_0} \delta \phi_0 \\ + \frac{\partial N_a}{\partial \lambda_0} \delta \lambda_0 + \frac{\partial N_a}{\partial a} \delta a + \frac{\partial N_a}{\partial f} \delta f \quad (25) \\ = \sum_i c_{a_i} \left( \Delta g_i + \delta \Delta g_i + \frac{\partial \Delta g_i}{\partial \gamma_e} \delta \gamma_e + \frac{\partial \Delta g_i}{\partial f} \delta f \right). \end{aligned}$$

This adjustment would be preferable to adjustment of the deflections because the variances of the astro-geodetic geoid heights will be appreciably smaller, and their values easier to estimate. The variances of the  $\Delta N$ , the differences in  $N$ , must first be found by methods such as are described in Ölander [1954] or estimated from

the misclosures in adjustment of the astrogeodetic net. Then each  $N_i = N_0 + \sum_{j=1}^i \Delta N_j$ . In matrix notation,  $N = FM$ , where the elements of  $N$  are the  $N_i - N_0$ ; of  $F$ ,  $+1, -1, 0$ , or fractions where more than two routes are used; and of  $M$ , the  $\Delta N_i$ . The  $N_i$  will thus have appreciable covariances, since  $V_N = FV_{\Delta N}F^*$ .

Geoidal heights adjustment could also be done as a two-stage solution analogous to that described in (16)-(20); that is, adjust properly weighted  $\delta N_{GR}$  and determine the  $\delta \Delta g_i$  therefrom.

Heiskanen [1957a] proposed minimizing the discrepancies between the astro-geodetic and gravimetric geoids to determine ellipsoid and datum parameters. (That is, minimize  $\sum (\xi_{GR} - \xi_A)^2$ ,  $\sum (\eta_{GR} - \eta_A)^2$ , and  $\sum (N_{GR} - N_A)^2$ .) Ledersteger [1951] utilized the first two of these conditions in his adjustment, and Lieberman [1955] and Fischer [1957b] used the last; all of these did not use weighting. This method also has the advantage of using considerable data already prepared. It entails about the same amount of work as the first stage of the method described in the previous paragraph, since there is still the labor of computing weights if the same accuracy is to be attained. The variance of each difference will be the sum of the astronomic and gravimetric variances,  $V_{A-GR} = V_A + V_{GR}$ . In the case of the astrogeodetic geoid elevations and all gravimetrically computed values, this matrix addition involves covariances as well. The weight matrix will, as usual, be the inverse of the variance matrix,  $P_{A-GR} = V_{A-GR}^{-1}$ . This method will not give directly an agreement between the gravimetric and astrogeodetic geoids, but it could be accomplished as follows:

(1) Assume the discrepancies  $(\xi_{GR} - \xi_A)$ , etc., remaining after the adjustment, to be eliminated by corrections in proportion to the weights; that is,

$$\begin{aligned} (\xi_{GR} - \xi_A) - \frac{\partial \xi_A}{\partial f} \delta f - \frac{\partial \xi_A}{\partial a} \delta a \\ - \frac{\partial \xi_A}{\partial \phi_0} \delta \phi_0 - \frac{\partial \xi_A}{\partial \lambda_0} \delta \lambda_0 - \frac{\partial \xi_A}{\partial N_0} \delta N_0 \\ = \frac{\rho_{GR}}{\rho_{GR} + \rho_A} \delta \xi_A - \frac{\rho_A}{\rho_{GR} + \rho_A} \delta \xi_{GR}. \quad (26) \end{aligned}$$

This would give agreement of geoids.

(2) Adjust the  $\delta \Delta g_i$  and  $\delta \gamma_e$  to the  $\delta \xi$ ,  $\delta \eta$ ,  $\delta N$

TABLE 1—Summary of methods

Method	Constants Weight = $\infty$	Observations $0 < \text{Weight} < \infty$	Parameters Weight = 0	Ignored
Area or arc	$\phi, \lambda, \Delta\phi, \Delta\lambda$	$\xi, \eta$	$a, f, \phi_0, \lambda_0, N_0$	$N, \Delta g, \gamma_e$
Geoidal heights	$\phi, \lambda, \Delta\phi, \Delta\lambda$	$N$	$a, f, \phi_0, \lambda_0, N_0$	$\xi, \eta, \Delta g, \gamma_e$
Gravity formula	$g$	$\Delta g$	$f, \gamma_e$	$\phi, \lambda, \text{etc.}$
Ideal		$\delta\phi, \delta\lambda, \delta\Delta\phi, \delta\Delta\lambda, \delta\Delta g, \delta\xi_L, \delta\eta_L$ ; sometimes $f$ .	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	
$\delta N$ neglected		$\delta\phi, \delta\lambda, \delta\Delta\phi, \delta\Delta\lambda, \delta\Delta y, \delta\xi_L, \delta\eta_L$ ; sometimes $\delta f$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N$
Geodetic fixed	$\Delta\phi, \Delta\lambda$	$\delta\phi, \delta\lambda, \delta\Delta g, \delta\xi_L, \delta\eta_L$ ; sometimes $\delta a, \delta f$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N$
Astronomic fixed	$\phi, \lambda, \Delta\phi, \Delta\lambda$	$\delta\Delta g, \delta\xi_L, \delta\eta_L$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N$
Reduced astro.	$\Delta\phi, \Delta\lambda$ ; some $\phi, \lambda$	$\delta\Delta g, \delta\xi_L, \delta\eta_L$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N$ ; some $\phi, \lambda$
Interpolated deflections	$\phi, \lambda, \Delta\phi, \Delta\lambda$	$\delta\xi, \delta\eta, \delta\Delta g, \delta\xi_L, \delta\eta_L$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N$
Interpolated geoid heights	$\phi, \lambda, \Delta\phi, \Delta\lambda$	$\delta N, \delta\Delta g, \delta N_L$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta\xi, \delta\eta$
Interpolated geoid heights	$\phi, \lambda, \Delta\phi, \Delta\lambda, N$	$\delta\Delta g, \delta N_L$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta\xi, \delta\eta$
Deflection difference	$\Delta\phi, \Delta\lambda$	$(\sum_i c' \cos \alpha \Delta g_i - \xi),$ $(\sum_i c' \sin \alpha \Delta g_i - \eta)$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta N, \delta\Delta g$
Geoid height difference	$\Delta\phi, \Delta\lambda$	$(\sum_i c \Delta g_i - N)$	$a, f, \phi_0, \lambda_0, N_0, \gamma_e$	$\delta\xi, \delta\eta, \delta\Delta g$

and the harmonic conditions as previously described to obtain anomalies consistent with the adjusted geoid.

*Discussion*—To facilitate comparison, the different methods of adjustment have been tabulated. In the observation column of Table 1 the quantity whose square is used in the square sum minimized is indicated; sometimes it is the 'observation' itself, for example,  $\Delta g$  or  $\xi$ ; sometimes it is a correction to the observation, for example,  $\delta\Delta g$  or  $\delta\phi$ . As previously noted,  $\Delta h$  is omitted throughout, and  $\delta N$  is indicated as "ignored" where appropriate, as has tacitly been done following Eq. (22). Also "constants" is perhaps the wrong term to use for some of the items in the weight =  $\infty$  column, since some of these will vary with change in the model and datum parameters. We have not distinguished the two-stage solution for the ideal method, since it is a mathematical transformation rather than a geodetically different method.

This paper represents a theoretical investigation of application of gravity data through Stokes' function to improve on the astro-geodetic geoid determinations made at the Army Map Service. For practical rather than for theoretical reasons, the application of gravity data at the Map Service is being made using spherical harmonics rather than Stokes' functions. However, the principal conclusions remain unchanged; the greatest problem is the statistical evaluation

of the gravity data to allow for its correlations and its proper weighting relative to the astro-geodetic data.

REFERENCES

- ARLEY, N. AND K. BUCH, *Probability and statistics*, John Wiley and Sons, 236 pp., 1950.
- BOMFORD, G. *Geodesy*, Oxford Univ. Press, 452 pp., 1952.
- BOMFORD, G., European geoid, *Report for 1954-57, Study Group 10, Sec. V, to I.A.G.*, XI Gen. Assembly IUGG, Toronto, 1957.
- CHOVITZ, B. AND I. FISCHER, A new determination of the figure of the Earth from arcs, *Trans. Amer. Geophys. Union*, **37**, 534-545, 1956.
- FISCHER, I., *A map of the geoidal contours in North America*, Army Map Service, presentation I.A.G., XI Gen. Assembly IUGG, Toronto, 1957a.
- FISCHER, I., *The Hough ellipsoid, or The figure of the Earth from geoidal heights*, Army Map Service, presentation I.A.G., XI Gen. Assembly IUGG, Toronto, 1957b.
- HEISKANEN, W. A., Discussion of "A new determination of the figure of the Earth from arcs", *Trans. Amer. Geophys. Union*, **38**, 579-580, 1957a.
- HEISKANEN, W. A., The Columbus geoid, *Trans. Amer. Geophys. Union*, **38**, 841-848, 1957b.
- HIRVONEN, R. A., The removal of spherical harmonics of first order from a field of observed gravity anomalies, *Pub. Fin. Geod. Inst.* **46**, pp. 59-68, 1955.
- HIRVONEN, R. A., On the precision of the gravimetric determination of the geoid, *Trans. Amer. Geophys. Union*, **37**, 1-8, 1956.
- JEFFREYS, H., The figures of the Earth and the Moon (third paper), *Mon. Not. R. Astron. Soc., Geophys. Sup.*, **5**, 219-247, 1948.

- KAULA, W. M., Accuracy of gravimetrically computed deflections of the vertical, *Trans. Amer. Geophys. Union*, **38**, 297-305, 1957.
- LAMBERT, W. D., Inadmissible spherical harmonics in the expansion of gravity anomalies, *Festschrift C. F. Baeschlin*, pp. 149-154, Zürich, 1957.
- LEDERSTEGE, K., Die Bestimmung des mittleren Erdellipsoides und der absoluten Lage der Landestriangulation, *Öster. Zs. Vermess.*, Sonderheft 12, 130 pp., 1951.
- LIEBERMAN, H. A., An investigation of the geoid in Europe, *Bul. Géod.*, no. 37, pp. 1-11, 1955.
- MOLODENSKIY, M. S., Bestimmung der Gestalt des Geoids unter gemeinsamer Anwendung astronomisch-geodätischer Lotabweich- und Schwerestörungen, *Verhand. Balt. Geod. Kom.*, pp. 203-223, 1937.
- MORELLI, C., Absolute determinations of gravity, connections between absolute stations and between first order stations, International formula for gravity, *Rep. Spec. Study Group 5, Section IV, I.A.G.*, XI Gen. Assembly IUGG, Toronto, 1957.
- O'KEEFE, J. A., AND J. P. ANDERSON, The Earth's equatorial radius and the distance of the Moon, *Astr. J.*, **57**, 219-247, 1952.
- ÖLANDER, V. R., On the weight function in astronomical levelling with areally distributed deviation-of-the-vertical stations, *Bul. Géod.* no. 34, pp. 329-342, 1954.
- RICE, D. A., Deflections of the vertical from gravity anomalies, *Bul. Géod.*, no. 25, pp. 285-308, 1952.
- TANNI, L., On the continental undulations of the geoid as determined from the present gravity material, *Pub. Fin. Geod. Inst.*, **18**, 78 pp., 1948.
- TENGSTRÖM, E., Outlines of a method for determining the geoid in Sweden by free-air anomalies, *Rik. All. Kart., Medd. 22*, 68 pp., Stockholm, 1954.
- TIENSTRA, J. M., An extension of the technique of the methods of least squares to correlated observations, *Bul. Géod.*, no. 6, pp. 301-335, 1947.
- UOTILA, U. A., Determination of the shape of the geoid, *Pub. Inst. Geod., Photogr., & Cart.*, **7**, Ohio State Univ., pp. 90-97, 1957.
- YAPLEE, B. S., R. H. BRUTON, K. J. CRAIG, AND N. G. ROMAN, Radar echoes from the moon at a wavelength of 10 cm, *Proc. I. R. E.*, **46**, pp. 293-297, 1958.
- ZHONGOLOVICH, I. D., On the determination of the dimensions of a general Earth ellipsoid, (in Russian), *Akad. Nauk. USSR, Inst. Teor. Astr., Trudy 6*, pp. 3-66, 1956.

(Communicated manuscript received February 25, 1958, and, as revised, July 24, 1958.)





## A Tentative World Datum from Geoidal Heights Based on the Hough Ellipsoid and the Columbus Geoid

IRENE FISCHER

*U. S. Army Map Service,  
Washington 25, D. C.*

**Abstract**—From all astro-geodetic material available at present, geoidal heights were computed on the 1927 North American Datum for the western hemisphere from Canada to Chile, and on the European Datum for the eastern hemisphere from Great Britain to Japan and from Scandinavia to South Africa. Improved reference ellipsoids were determined under various conditions for each hemisphere by the least squares method. Triaxiality is refuted. The Columbus Geoid, gravimetrically derived by W. A. Heiskanen, was used to connect the astro-geodetic systems of the two hemispheres. By minimizing the differences between the astro-geodetic and gravimetric geoid elevations at 75 points in the western hemisphere and 127 points in the eastern hemisphere, a theoretically absolute orientation based on the Hough ellipsoid was determined, leading to a tentative World Datum.

The use of geoidal heights rather than deflections of the vertical for a determination of the figure of the earth depends on the availability of geoidal contour maps for large contiguous areas. In Figure 1 the area covered is shown by hatching; the heavy lines indicate the long arcs of our preliminary study [Chovitz and Fischer, 1956], where the method of deflections of the vertical had been employed.

*The western hemisphere*—A geoidal contour map of North America on the 1927 North American Datum was completed recently [Fischer, 1957a]. Figure 2 is a generalization of this map, taking an initial value of zero at Meades Ranch instead of the 10 m for Calais, Maine, which had been used by Fischer [1957a] in conformity with Hayford's [1909] map. A geoidal profile was carried along the first-order triangulation arc through Central and South America with two eastward spurs, one through Bolivia into Brazil, the other through Colombia and Venezuela to Trinidad. In Trinidad this arc connects with the Hiran trilateration coming from Florida, closing the loop of about 10,000 km around the Caribbean Sea with a discrepancy of less than 25 m in position.

Figure 3 shows the geoidal profile on the 1927 North American Datum along the 100th meridian in North America and along the triangulation arc through Central into South America. The geoidal heights were computed according to the projection method, in which distances are reduced to the spheroid, not simply to the geoid

as in the customary development method. Molodenskiy's [1944] technique was used to compute this additional correction (from geoid to spheroid), which is essentially a scale-factor correction. It is described in Fischer [1957b] in more detail. The effect on the geoidal profile is illustrated in Figure 3 as the difference between the full line (projection method) and the upper broken line (development method); it is negligible in North America, but accumulates to about 60 m at 40° South. The lower broken line is taken from the preliminary study [Chovitz and Fischer, 1956] where this arc was treated as a meridian, neglecting the  $\eta$  component of the deflection of the vertical.

The geoidal height relative to a reference ellipsoid depends on the size and shape of that ellipsoid and on its orientation with respect to the geoid, altogether on five parameters. By varying these five parameters and minimizing the geoidal heights in a least-squares solution, a best-fitting ellipsoid can be determined. The well known formula by de Graaff Hunter [Survey of India, 1939] was used for this purpose. To represent the western hemisphere, 131 observation equations were formed for points at 5° intervals. The origin was taken at Meades Ranch. The parameters of the best-fitting ellipsoid were determined as  $f = 1/(297.1 \pm 0.2)$  and  $a = 6,378,239 \pm 16$  m [Fischer, 1957b]. The geoid contours of North America referred to this ellipsoid are plotted in Figure 4. A profile along the 100th meridian, continuing into the Central

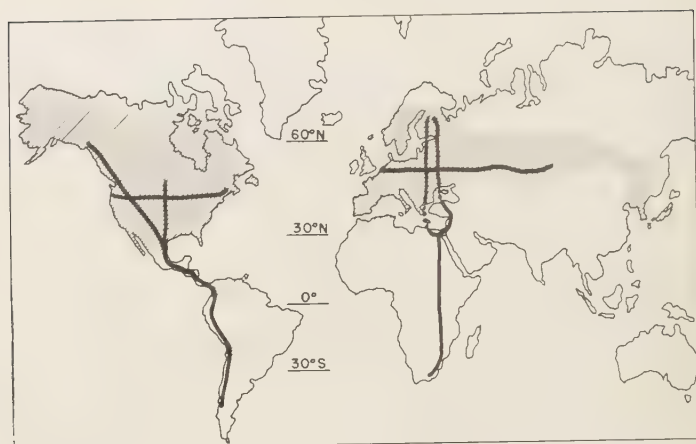


FIG. 1—Coverage

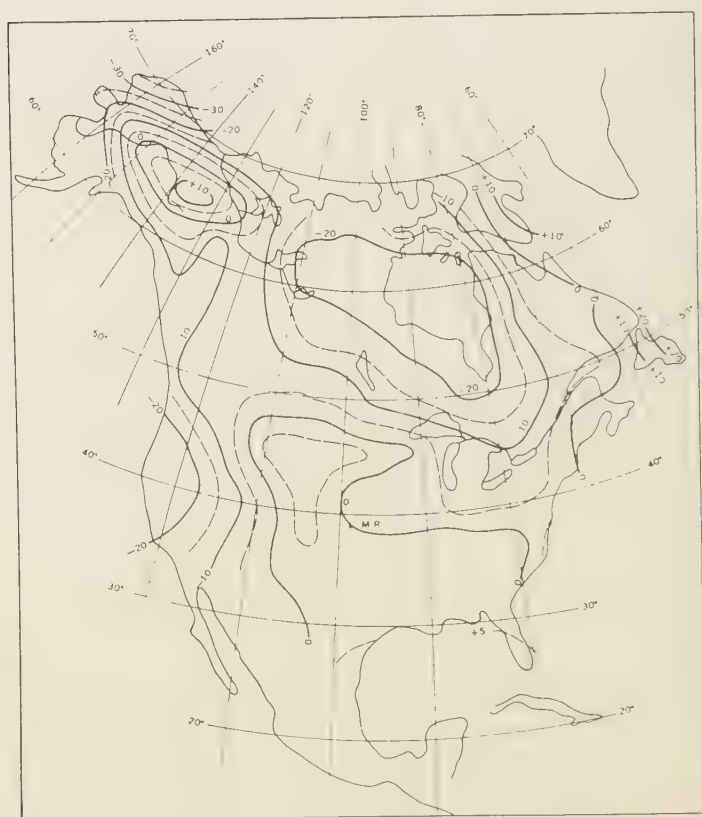


FIG. 2—North America, geoid contours in meters, 1927 North American Datum, assumed initial value zero meters at Meades Ranch (M.R.)

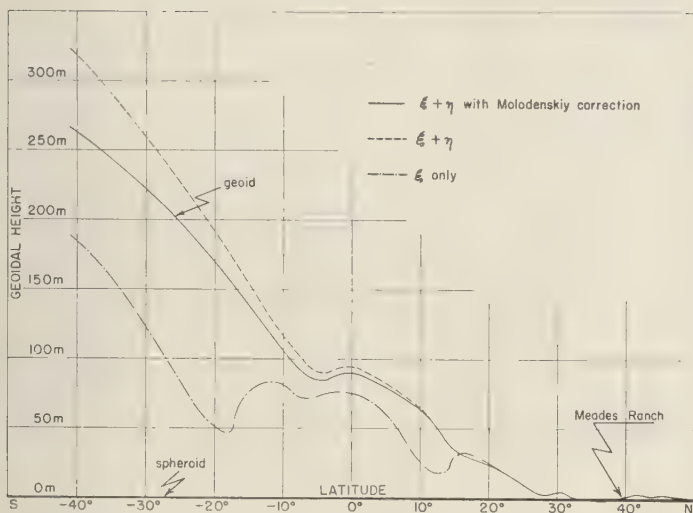
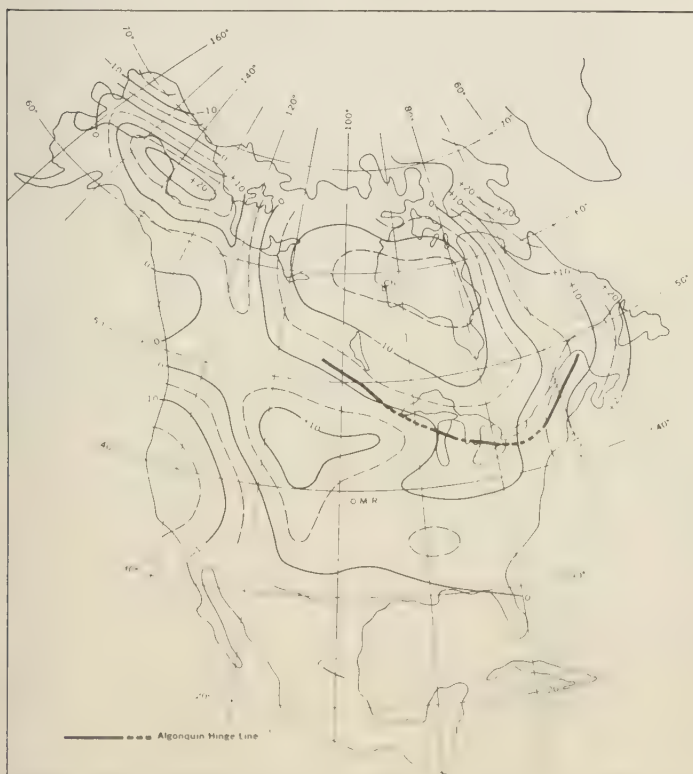


FIG. 3—Geoidal heights, America, 100th meridian, 1927 North American Datum





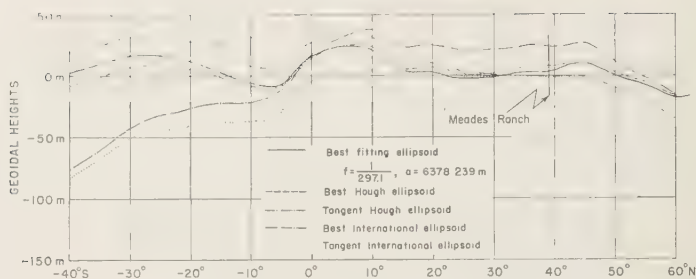


FIG. 5—Geoidal heights relative to various ellipsoids, America, 100th meridian

and South American arc, is plotted in Figure 5. This Figure also illustrates the difference in geoidal heights, if the best fitting ellipsoid were replaced by the International or the Hough ellipsoid, both in their best fitting orientation and in the case of tangency at the origin ( $\delta\xi_0 = \delta\eta_0 = N_0 = 0$ ). The zero line represents each of the ellipsoids; the symbol of the curved line, representing the corresponding geoid profile, indicates to which ellipsoid the geoidal heights are referred. It can be seen that the International ellipsoid does not fit well outside the United States.

The Hough ellipsoid ( $f = 1/297$ ,  $a = 6,378,270$  m) is a conventional surface used in problems of world-wide extent as in Project Vanguard.

*The eastern hemisphere*—The geoidal heights in Europe and North Africa on the European Datum were taken over from Lieberman's [1955] work. The addition of England, Scotland, and the Shetland Islands [Assn. Internat. Geod., 1957] provided a loop closure to Norway. The connection from Greece via Crete to Africa changed the height of Crete to a value larger than that on Lieberman's chart; the change was confirmed by a loop closure in Africa through the use of an alternative route via the Levant. The Molodenskiy correction is insignificant for Europe but is quite large in South Africa. Its effect along the 30th meridian in Africa is shown in Figure 6, together with the geoidal profile along this arc.

The geoidal contour chart for the USSR was given on the Bessel ellipsoid of the 1932 triangulation [Army Map Service, 1953]. This reference ellipsoid and datum were changed to the International ellipsoid and European Datum by computing the separation between the two

ellipsoids at stations covering the whole area at 5° intervals, making use of the known difference in position of Pulkovo. The underlying assumption here is that the Russian chart represents true geoidal heights according to the projection method. The Russian source gives no explanation on the method of computing these heights. The brief statement that they are based on triangulation data as of 1936 indicates that some correction should have been applied, since we know that at that time bases were not yet reduced to the ellipsoid as a general procedure. No mention of that correction having been made raised some doubt as to whether or not we should apply the Molodenskiy correction to the Russian material before converting it to the European Datum. In the paper presented at Toronto [Fischer, 1957b], both alternatives were pursued numerically. The amount of this correction is certainly not negligible, accumulating to about 25 m at the eastern end of the chart at 130°E, in case the original triangulation was computed on the Bessel ellipsoid. If a well fitting ellipsoid was used, the correction would be small. On the

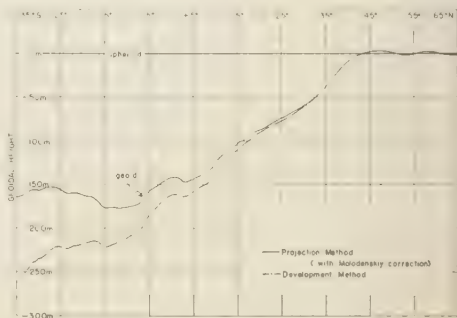


FIG. 6—Geoidal heights, Africa, European Datum, effect of Molodenskiy correction

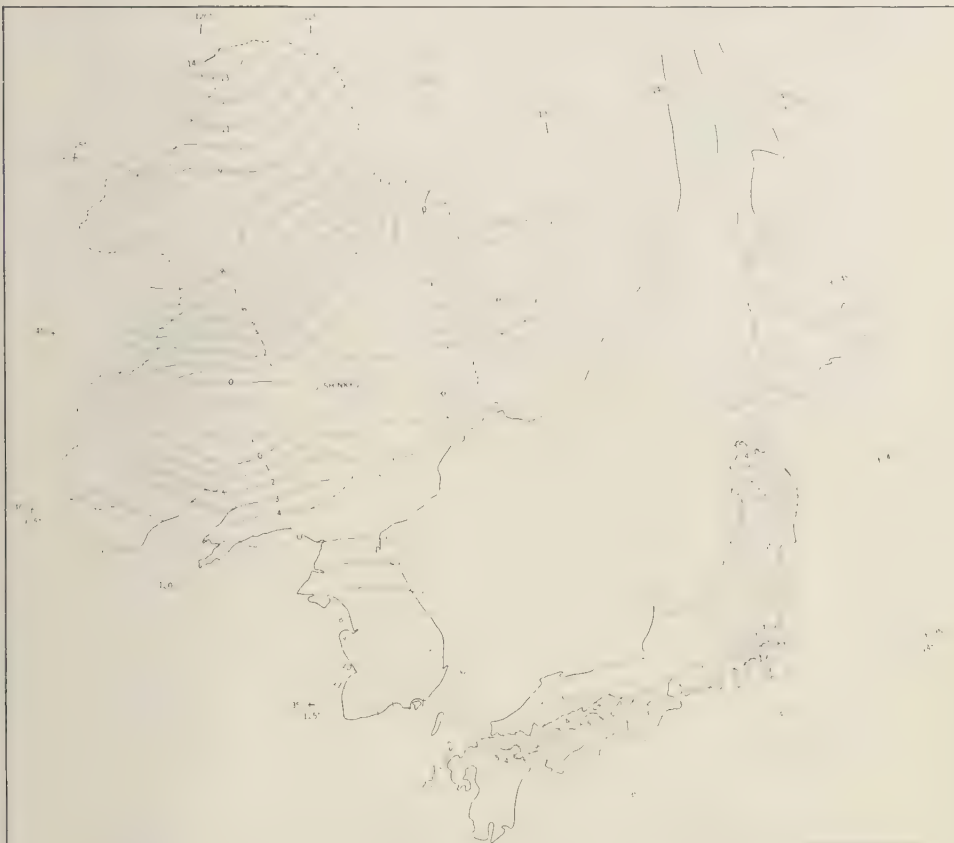


Fig. 7—Manchuria, Korea, Japan, geoidal contours in meters, Manchurian Principal System

other hand, we know from the historical background that at the time of Dubovskoy's publication in 1939 the distinction between development and projection methods was very well known and the implications had been discussed for more than a decade. Thus we feel justified in assuming for the present that these implications were considered when the chart was drawn. It is planned, however, to continue the alternative computations until the facts can be established.

The Manchurian material, given in the form of astro-geodetic deflections on the Bessel ellipsoid, was first used to compute geoidal contour lines on the Manchurian Datum by Hayford's [1909] method; charts were drawn showing lines of equal deflections for the meridional and prime vertical components. From these charts geoidal heights were derived (Fig. 7)

in the same way as for the North American map. No Molodenskiy correction seemed necessary here for the same reasons that held for Europe and North America. The Manchurian Datum could now be changed to the European Datum, since there is an overlap with the Russian net at  $50^{\circ}\text{N}$ ,  $130^{\circ}\text{E}$ .

The astro-geodetic deflections in Japan on the Tokyo Datum [Geogr. Surv. Inst., 1951, 1953, 1955; Okuda, 1951; Torao, 1951] were used in the same way to construct first lines of equal deflections, and from these, the geoidal contours of Figure 8. An extension of this chart was computed along a profile through Korea. A connection to Manchuria was computed, by which all of Japan and Korea could be converted into the Manchurian Principal System and from there into the European Datum. A com-



FIG. 8—Japan, geoidal contours in meters, Tokyo Datum

parison of the geoid contours of Japan on the Manchurian and Tokyo Datums (Figs. 7 and 8) reveals that the Manchurian Datum is much better suited to Japan than is its own datum. The reason is that Tokyo, the chosen origin, happens to be an atypical point, lying on a steep

geoidal slope towards the Ramapo Deep in the Japanese trench. The resulting absolute deflection must be quite large. Since zero deflections at Tokyo are assumed as the basis for the Tokyo Datum, the geoidal heights of the country must appear as systematically negative. The deflection

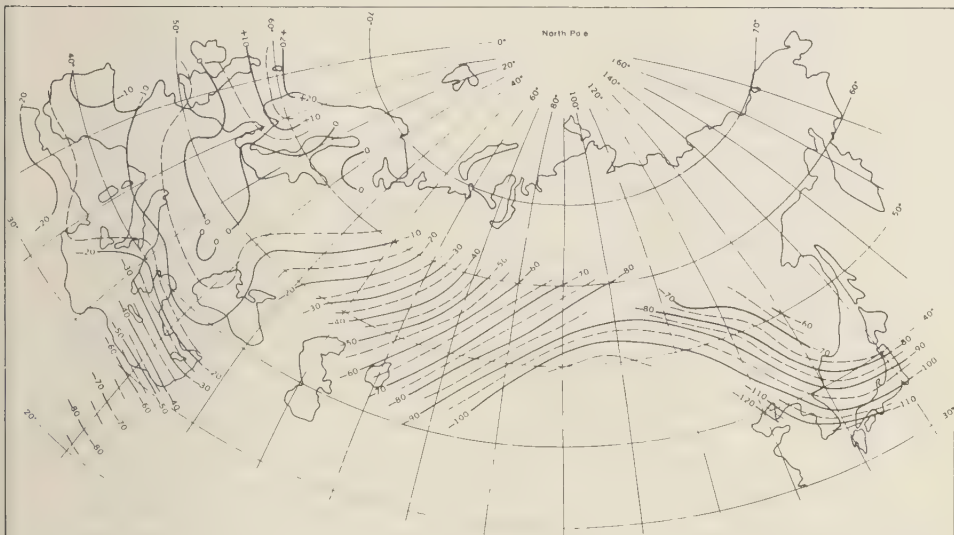


Fig. 9—Eastern hemisphere, geoid contours in meters, European Datum

of Tokyo on the Manchurian Principal System turns out to be of about the same order as, though somewhat larger than, the results found by *Atumi* [1933] and *Kawabata* [*Okuda*, 1951, p. 246] in their studies of a best-fitting ellipsoid.

The vast area from the Shetland Islands to Japan and from Scandinavia to South Africa is now unified on the European Datum (Figs. 6 and 9). The point of origin was chosen at 50°N, 25°E, and 137 equations were formed at 5° intervals. The parameters of the best-fitting ellipsoid are  $f = 1/(297.1 \pm 0.2)$  and  $a = 6,378,279 \pm 8$  m; these are somewhat larger than the figures given by *Fischer* [1957*b*] because new information has been obtained for Great Britain, South Africa, and Japan. The contours of the geoid referred to this ellipsoid are plotted in Figure 10. Figure 11 illustrates the differences caused by the use of other reference ellipsoids. As in the case of the western hemisphere, the Hough ellipsoid is a good approximation to the best solution, while the International ellipsoid, even in its best fitting orientation, does not fit well.

The Krasovskiy ellipsoid, oriented as in the present Russian datum, was also investigated as a reference spheroid. As expected, it fits very well along the parallels of the northern hemisphere where it was derived. It does not fit in Africa, or in the Far East.

*The Ice Age*—In Figures 4 and 10 a remarkable correlation appears between the geoid contours and the areas of Pleistocene glaciation. The connection between them is discussed in a separate article [*Fischer*, 1959].

*Triaxiality*—The representation of the geoid on a best-fitting ellipsoid should solve the problem of triaxiality. In Figure 10 the highest and lowest areas beyond the  $\pm 20$  m contour line are marked. They seem to be arranged in a distinct undulation pattern. The longitudinal distance between the high and low zones, however, is not 90° but only 45°. This would suggest a system of two major axes perpendicular to each other, with two minor axes midway between them; in other words, a pentaxial rather than a triaxial spheroid. A continuation of this undulation pattern toward the west could be seen in the hump one might expect on the ice cap of Greenland and in the North Atlantic, the depression of the Hudson Bay and the hump again in the area of about 135° west longitude. Before assuming an undulating spheroid, triaxial or otherwise, for the whole earth, it would have to be seen whether the same pattern holds within the entire longitude zone. Apparently, this is not the case. Therefore, the ellipsoid of revolution still seems to be our best choice of reference ellipsoid.

*The Columbus Geoid*—Gravimetrically computed geoidal heights are deviations from a



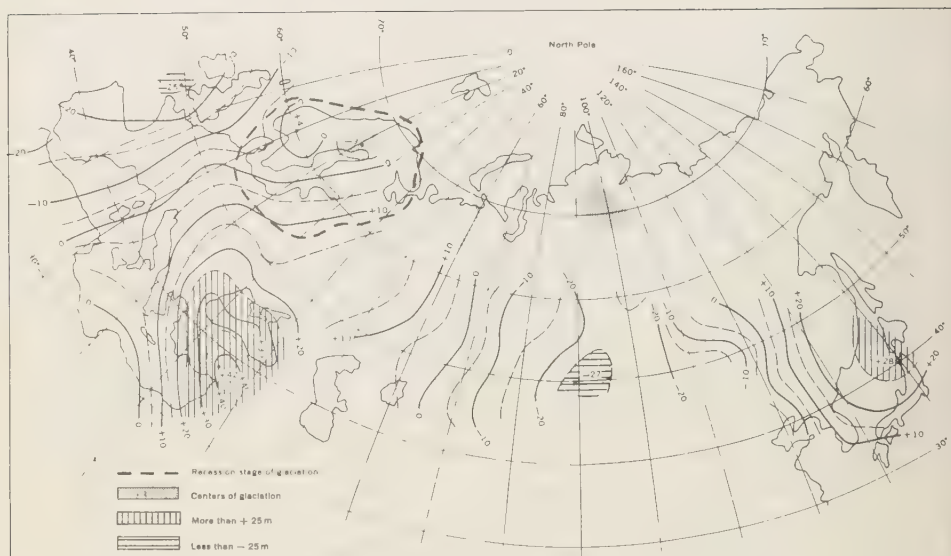


FIG. 10—Eastern hemisphere, geoid contours in meters, best fitting ellipsoid

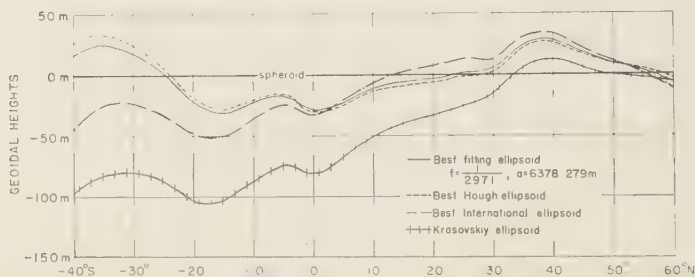


Fig. 11—Geoidal heights relative to various ellipsoids, Africa, 30th meridian

norm based on the adopted gravity formula. The Columbus Geoid [Heiskanen, 1957] is based on the International gravity formula in which the value of the flattening  $f = 1/297$ . The norm, therefore, is an ellipsoid of revolution (supposing the longitude term of that formula to be neglected) with a predetermined shape. The size of that ellipsoid or the distance of a particular point from the center of the earth cannot be determined by this method alone, if we are interested in an accuracy of the order of 50 m as is discussed in astro-geodetic studies. We might envisage the gravimetric geoid as a surface that, within reason, can be stretched to scale, without effect on the geoidal elevations; something like an elastic wrist watch band. Matching an

astro-geodetic to a gravimetric geoid of the same area imposes a definite scale on the latter. If the size of the gravimetric ellipsoid (the norm) is known, a conversion formula can be derived between the gravimetric and astro-geodetic systems.

The Columbus Geoid could be matched at 75 out of 131 points to the astro-geodetic geoid on the North American Datum. The least-squares solution for an ellipsoid of flattening  $f = 1/297$ , compatible with the Columbus Geoid elevations at these 75 points, gives the size of that ellipsoid as  $a = 6,378,240 \pm 21$  m. The comparable astro-geodetic ellipsoid, by which we mean the best fitting ellipsoid with the same flattening of  $1/297$  for the same 75 points, has a

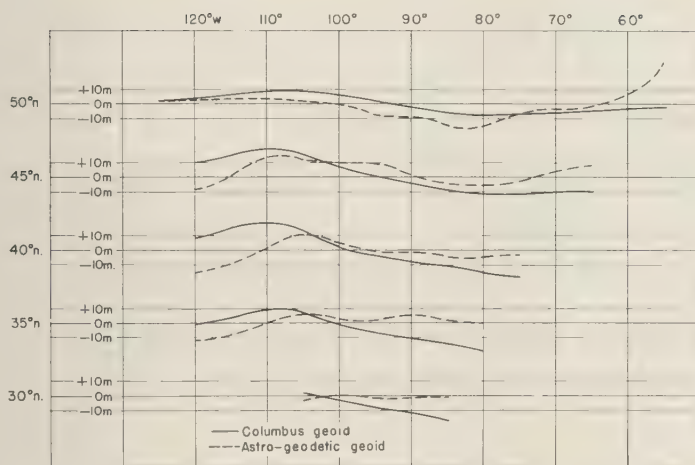


FIG. 12—Columbus Geoid in America, profiles along parallels

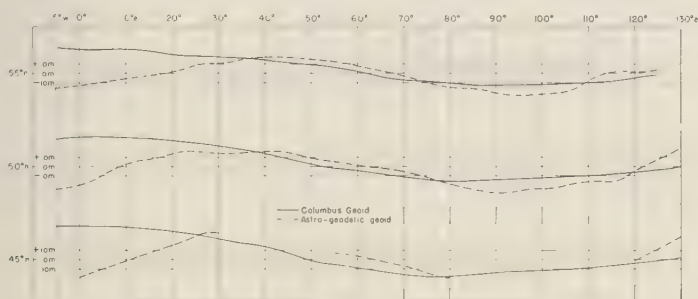


FIG. 13—Columbus Geoid in eastern hemisphere, profiles along parallels

size of  $a = 6,378,248 \pm 25$  m, practically the same size as the above gravimetric ellipsoid. This may be taken as a criterion for good agreement. Figure 12 illustrates this agreement, comparing profiles of the Columbus Geoid with those of the astro-geodetic geoid, referred to this best fitting ellipsoid (not to the North American Datum). It can be seen that a certain tilt would make the profiles practically coincide, if we disregard certain areas where obviously more information is needed.

The Columbus Geoid of Europe, Africa, and Asia was matched at 127 points to the astro-geodetic geoid of the eastern hemisphere derived here. In the least-squares solution for the size of the gravimetric ellipsoid the semi-major axis  $a$  was found to be  $6,378,214 \pm 8$  m. The comparable astro-geodetic ellipsoid has a size of  $a = 6,378,270 \pm 10$  m; it does not differ greatly

from the solution for all points, although the African points south of the equator and the northern points on the 65th parallel were omitted because they were not covered by the Columbus map. This time our criterion of good agreement has not been met, since the semi-major axes derived from the two geoid maps differ by more than 50 m. In Figure 13, the profiles along three parallels are compared. The astro-geodetic profiles of this Figure are again not referred to the European Datum but to the best fitting comparable ellipsoid. The Figure suggests that there is no single shift that could make both ends of the profiles coincide; the European parts could be made to fit at the expense of Asia; the Asiatic parts agree in a very generalized sense and that only at the expense of Europe. It is obvious that much more work is needed to find and remove the causes of disagreement. If the



astro-geodetic geoid is at fault, it might be a reflection on the weakly documented section in the USSR. If the gravimetric geoid is the one that should be changed it might be for the reason of insufficient observations in certain areas.

*A tentative World Datum*—It is said that the gravimetric system is absolutely oriented, meaning that the center of the ellipsoid used as the norm is located at the center of the earth and its minor axis coincides with the axis of rotation. The theoretical relation to the center of the earth is the big advantage of the gravimetrical approach over the astro-geodetic approach, for it relates otherwise unconnected areas in a single system. On the other hand, the size of this ellipsoid cannot be determined without recourse to the astro-geodetic method. Thus the two methods should be combined and the combination used to connect the North American Datum with the European Datum and to devise a World Datum.

Matching the gravimetric geoid elevations to astro-geodetic geoid heights in order to determine the size of the corresponding gravimetric ellipsoid has yielded different sizes for different areas. For the western hemisphere we found  $a = 6,378,240$  m, for the eastern hemisphere  $a = 6,378,214$  m; for Europe alone Lieberman [1955] determined Tanni's ellipsoid as  $a = 6,378,160$  m. This indicates that the gravimetric ellipsoid depends as much on the particular area as does the astro-geodetic ellipsoid. Our astro-geodetic studies have led to the conclusion that the Hough ellipsoid is our present best guess for the figure of the earth as a whole. We will, therefore, assume this size also for the gravimetric norm. A least-squares solution for the 75 points in the western and 127 points in the eastern hemisphere, minimizing the differences between astro-geodetic and gravimetric geoid heights, determines the best orientation of the North American and European systems in relation to the so-called gravimetric Hough ellipsoid and leads to conversion formulas between them.

The tentative World Datum is defined by the parameters of the Hough ellipsoid ( $a = 6,378,270$  m,  $f = 1/297$ ) and the World geodetic coordinates of Meades Ranch, or any other point chosen as origin. Figure 14 is a geoid contour map based on this World Datum. The elevations shown are essentially astro-geodetic elevations. Only the

orientation was determined from gravimetric information and used to convert to the World Datum the elevations previously determined on the European or North American Datum respectively. While theoretically the astro-geodetic and gravimetric geoid elevations should agree, this World Geoid map and the Columbus Geoid map do not agree in detail at this time.

The World Datum defined here is as tentative as are the Hough ellipsoid and the Columbus geoid, on which it is based. It must be borne in mind that only a small fraction of the earth's surface is covered at present by astro-geodetic or gravimetric data. Future information will certainly modify our numerical results, especially information obtained from two projects: the North Atlantic tie between the hemispheres and the artificial satellite.

*Acknowledgments*—Appreciation is expressed to Mary Slutsky, Army Map Service, for assistance in the computations, and to John A. O'Keefe, Chief of Research and Analysis Branch, Army Map Service, for his generous support and encouragement.

#### REFERENCES

- ARMY MAP SERVICE, Development of a method of utilizing gravimetric surveys for cartographic-geodetic networks, *Techn. Rep. 12*, 16 pp., 1953.
- ASSOCIATION INTERNATIONALE DE GÉODÉSIE, *Liste des stations de déviation de la verticale rattachées au Réseau Européen*, 1957.
- ATUMI, K., La déviation de la verticale au Japon, *Rikuchi-Sokuryobu* (Imperial Land Survey), 9 pp., 1933.
- CHOVITZ, B. H., and I. FISCHER, A new determination of the figure of the Earth from arcs, *Trans. Amer. Geophys. Union*, **37**, 534-545, 1956.
- FISCHER, I., A map of geoidal contours in North America, presented to the 11th General Assembly of the IUGG, Toronto, 1957a (available at Army Map Service).
- FISCHER, I., The Hough ellipsoid or the figure of the Earth from geoidal heights, presented to the 11th General Assembly of the IUGG, Toronto, 1957b (available at Army Map Service).
- FISCHER, I., The impact of the Ice Age on the present form of the geoid, *Journal of Geophysical Research*, **64**, 85-871, 1959.
- GEOGRAPHICAL SURVEY INSTITUTE, The observation of the vertical deflection in Japan, *Ministry of Construction, Japan, Bul. Geogr. Surv. Inst.*, **2** pt. 2-3, 167-237, 1951; **3**, pt. 2-4, 149-232, 1953; **4**, pt. 3-4, 15-161, 1955.
- HAYFORD, J. F., The figure of the Earth and isostasy from measurements in the United States, *Coast and Geod. Surv.*, s. p. 82, 178 pp., 1909.
- HEISKANEN, W. A., The Columbus Geoid, *Trans. Amer. Geophys. Union*, **38**, 841-848, 1957.



- LIEBERMAN, H. A., An investigation of the geoid in Europe, *Bul. Géod.*, **37**, 1-11, 1955.
- MOLODENSKIY, M. S., Principal problems in astrometric-gravimetric levelling of large areas, *Sbornik Nauchno-Tekhnicheskikh i Proizvodstvennykh Statei po Geodezii, Kartografii, Topografii, Aero-s'emke i Gravimetrii*, no. 4, 11 pp., Moscow, 1944 (in Russian).
- OKUDA, T., On the change of local geoid in the southwestern part of Japan, *Ministry of Construction, Japan, Bul. Geogr. Surv. Inst.*, **2**, pt. 4, 239-275, 1951.
- SURVEY OF INDIA, *Professional Paper No. 28*, p. 40, 1939.
- TORAO, M., Deflections of the vertical in Japan, *Jap. Geod. Comm.*, Rep. 1, n.s., 34 pp., 1951.
- (Manuscript received July 7, 1958; presented at the Thirty-Ninth Annual Meeting, Washington, D. C., May 5, 1958.)

# The Impact of the Ice Age on the Present Form of the Geoid

IRENE FISCHER

U. S. Army Map Service,  
Washington 25, D. C.

*Abstract*—The shape of the geoid in North America and Fennoscandia shows a remarkable correlation with the extent of Pleistocene glaciation in these areas.

The modern specialization of scientific effort leaves us often unaware of new findings and theories in other fields, findings which may corroborate some aspects of our own research. Whether there is outright confirmation of results or only tentative correlation, the very difference in method and purpose of the specialized fields of interest provides a certain depth of insight and a challenge for further investigations.

One of the main objects of geodesy is the determination of the shape of the earth. Geographers and geomorphologists describe the topography: the mountains and valleys, continents and oceans. Geodesists are more interested in measuring the earth as a whole and regard the topographic features as local deviations from the sea-level surface, the geoid. In the course of time our concept of the shape of the earth or, more precisely, the geoid, has developed from a plane to a sphere, then to an ellipsoid, and finally to an irregular surface whose exact features are as yet unknown over wide areas. The way to cope with such an irregular surface mathematically, as is the obligation of the geodesist, is to find a regular surface closely approximating the geoid, such as a suitably-shaped ellipsoid, and measure the deviations from point to point. Thus our task is divided into two parts: (1) to find the dimensions of the best ellipsoid to serve as an approximation, and (2) to find the deviations of the geoid from this ellipsoid in specific areas of the world. These deviations are called geoidal heights. For areas where we know sufficiently many of such heights we can draw geoidal contour maps in exactly the same manner as topographic contour maps (hypsometric maps) are drawn. In the one case the irregular surface of the geoid is referred to an artificial, but highly useful, regular surface as basis, in the other case the irregular topography is referred to sea level.

Recently, Army Map Service made a geoidal contour map of North America [Fischer, 1957a] and a relief model of this part of the geoid [Dille, 1958]. A generalized geoid map of the same area appears in another paper by the author in this issue as Figure 4, p. 75. Figure 10, (p. 80) of that paper represents a map of the geoid in the eastern hemisphere [Fischer, 1957b].

From the geophysical viewpoint, the irregularities of the geoid are an expression of irregularities of the masses in the earth's crust. The level in the surveyor's theodolite is affected, for example, by the masses of a nearby mountain or by heavy subterranean masses, so that the level surface (the geoid) shows a hump in that region; if there is a deficiency of mass in a certain region compared with the surrounding areas, the geoid shows a depression. Why certain regions have a relative excess or deficiency of mass is still a matter for speculation. This is one of the problems for which a combination of efforts from various special fields of study should be most fruitful. An attempt in that direction is made here.

Figure 1 shows the extent of the Pleistocene glaciation in North America and the direction of the flow of ice [Flint, 1952]. The huge Laurentide ice sheet centered on present-day Hudson Bay and extended beyond the Great Lakes and the NW-SE series of smaller lakes, all of which are residuals of a recessional stage. Other centers of glaciation were in the western mountains and in Greenland.

It is assumed that the weight of these huge ice masses made a depression in the crust of the earth [Daly, 1934]; when the ice melted and the weight diminished the process was reversed, resulting in a so-called postglacial uplift which is gradually eliminating the depression in a centripetal direction. The amount of this uplift is estimated at various places to be several

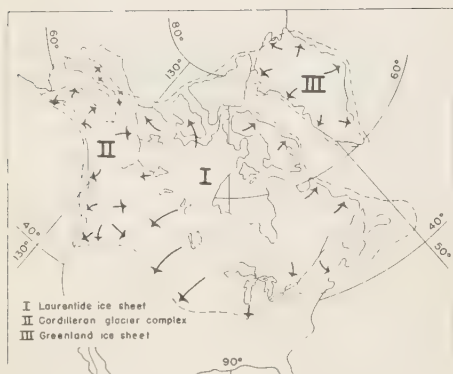


FIG. 1—Pleistocene glaciation of North America, after Flint [1952]

hundred meters, and it is expected to increase at the rate of about one to three meters per century. The observations on which these estimates are based refer primarily to marine and lacustrine shorelines high above present water levels, and to other evidence such as traces of human settlements, or the rate of decay in driftwood strand lines. In some areas, painstaking investigations were made of tilted water planes by studying the angles of sediment layers [Parmenter, 1956] or by checking water gages with repeated leveling [Renquist, 1952]. The reconstruction of the history of postglacial uplift [Lougee, 1953] points to several intermittent stages bounded by so-called hinge lines where the angle of tilt changes. These hinges are roughly concentric breaks in the earth's crust, parallel to the limits of the glaciation and its recessional stages.

If Figure 4, above cited, and Figure 1 are compared in the light of the foregoing, the geoidal depressions in the Hudson Bay region, in the area on the West Coast, and in Alaska assume peculiar significance. The center of depression in the Hudson Bay region is near Churchill, where the rate of uplift has been estimated to be three meters per century. The Algonquin hinge line [Lougee, 1953], marking the boundary of the area where uplift still occurs, roughly parallels our geoidal contour lines. There is a north-south ridge between this depression and the one in the west. This ridge parallels the area between opposing arrows on the glaciation chart indicating a boundary

zone between the two glaciers. Similarly, in Alaska, the Yukon valley, which was not glaciated, appears as a geoidal ridge between the once heavily glaciated mountains in the north and in the south. The Greenland glacier was much greater than it is today; the zone where its ice flow encountered the one from the Laurentian glacier should then appear as a geoidal ridge. Our geoidal chart does not as yet include that region, but the geoid definitely rises in that direction.

In Europe, the correlation between geoid contours and Pleistocene glaciers is less striking, partly because these were much smaller. Figure 10, cited above, shows the main centers in Fennoscandia and Siberia [Thiel, 1951]. Our geoidal chart does not reach the areas of the Siberian glaciers but it does show a depression in Sweden, on the west coast of the Gulf of Bothnia, at the center of glaciation. The limits of a recessional stage are shown [Gottmann, 1954]. In the east and southeast this line parallels the geoidal contour lines. The confluence between the Fennoscandian ice sheet and the glaciers from the Ural, appears again as a geoidal ridge. In the south and west, however, there is no correlation as the geoidal contour lines suddenly veer to the south. According to Ewing and Donn's [1956] theory, the Arctic Ocean must have been ice free at times of glaciation. The open ocean caused precipitation over the adjacent cold land and thus provided nourishment for the glacier. Nourishment from the south by storms may have been hampered by the mountain ranges. This would explain why the Scandinavian ice sheet was less thick towards the south than towards the north. This also might explain why the impact of the ice load on the form of the geoid is not discernible towards the south.

*Acknowledgment*—I am indebted to John A. O'Keefe, Chief of Research and Analysis Branch, Army Map Service, for the challenging suggestion which led to this study.

#### REFERENCES

- DALY, R. A., *The changing world of the ice age*, Yale University Press, New Haven, 1934.  
 DILLE, J., The missile era race to chart the Earth, *Life*, 44, 130, May 12, 1958.  
 EWING, M., AND W. L. DONN, A theory of Ice Ages, *Science*, 123, 1061-1066, 1956.  
 FISCHER, I., A map of geoidal contours in North America, presented to the 11th General Assembly

of the IUGG, Toronto, 1957a (available at Army Map Service).

FISCHER, I., The Hough Ellipsoid, or the Figure of the Earth from geoidal heights, presented to the 11th General Assembly of the IUGG, Toronto, 1957b (available at Army Map Service).

FLINT, R. F., The Ice Age in the North American Arctic, *Arctic*, 5, p. 140, 1952.

GOTTMANN, J., *Geography of Europe*, Holt and Co., New York, p. 130, rev. ed., 1954.

LOUGEE, R. J., A chronology of postglacial time in Eastern North America, *The Scientific Monthly*, 76, 259-276, 1953.

PARMENTER, G., Glacial water levels in Narragansett Basin and Blackstone River Valley, manuscript available at Clarke University Library, 111 pp. 1956.

RENQUIST, H. Endogenetic phenomena, *Suomi, General handbook on the geography of Finland*, Geogr. Soc. of Finland, 36-48, 1952.

THIEL, E., Die Eiszeit in Sibirien, *Erdkunde*, Bd. V, 16-35, 1951.

(Manuscript received July 7, 1958; presented at the Thirty-Ninth Annual Meeting, Washington, D. C., May 5, 1958.)





## A Method of Evaluating the Effect of a Monomolecular Film in Suppressing Reservoir Evaporation

G. EARL HARBECK, JR., AND GORDON E. KOBERG

*U. S. Geological Survey, Denver, Colorado*

**Abstract**—Evaporation from a water surface may be reduced by the application of a monomolecular film. Because of possible seasonal variations in evaporation-pan coefficients, computed savings in evaporation may be unreliable if based upon evaporation-pan records. A more reliable measure of the evaporation reduction is obtained by the solution of equations based upon energy budget and heat and mass transfer theory. The rise in water-surface temperature caused by application of the film can be computed.

In recent years the possible use of a monomolecular film to suppress reservoir evaporation has received considerable attention. Annual gross evaporation from Lake Mead, the Nation's largest reservoir, ranged from 699,000 to 875,200 ac ft during the water years 1953–1956. The amount of water lost by evaporation from this one reservoir is greater than the total capacity of most reservoirs in the United States. Any practical method of reducing the loss from Lake Mead and other reservoirs in the arid West would be invaluable. Not only is the lost water of considerable intrinsic value, it is also pure water. It is possible that some water supplies now unusable because of slightly excessive amounts of dissolved solids might become usable if evaporation were suppressed.

*Mansfield* [1953] experimented with monomolecular films for the suppression of evaporation in Australia in 1952–1953. Although his first experiments were unsuccessful, he realized the potential value of a practical technique and continued his investigations. Prior to his work, other workers, including *Rideal* [1925], *Langmuir* and *Langmuir* [1927], and *Langmuir* and *Schaefer* [1943], found that monolayer films diminished the rate of evaporation, but apparently no attempts were made to apply the technique to reservoirs.

*Mansfield* [1956] recommended the use of an evaporimeter to assess the results of the field trials of hexadecanol. His method consisted of measuring changes in reservoir stage during a pre-treatment calibration period. He separated the observed stage changes into the two components of seepage and evaporation by comparison with an evaporimeter record and determined a 'pan coefficient' for the evaporimeter. He

assumed that the seepage loss and pan coefficient thus determined during the calibration period were applicable after the film had been applied. After treatment, he measured changes in stage in the reservoir and adjusted them for seepage to obtain the actual evaporation loss. He determined the evaporation that would have occurred had the film not been applied by using the evaporimeter data multiplied by the pan coefficient.

The assumption that the amount of seepage loss was not affected by the application of the film appears reasonable, provided there was no great change in reservoir stage.

The assumption that the pan or evaporimeter coefficient determined during the calibration period would be applicable after treatment is not tenable. The Australian evaporimeters referred to by *Mansfield* are sunken pans. The Lake Hefner studies [*Kohler*, 1954] included tests of sunken pans. The Colorado sunken pan was three feet square and 18 inches deep; the Bureau of Plant Industry (BPI) pan was six feet in diameter and two feet deep. One type of Australian evaporimeter is a sunken tank three feet in diameter and three feet deep, surrounded by an outside tank four feet in diameter. The annular space between the two tanks is filled with water to minimize heat exchange between the inner pan and the ground. *Kohler's* data indicated that the range in monthly pan coefficients for the Lake Hefner sunken pans was greater than for the above-ground Class A pan. For the sunken pans, values of the monthly pan coefficient ranged from about 0.2 to 1.5. The monthly variation in the coefficient to be used with the above-described Australian evaporimeter may be less than this because of the insulating water

jacket, but substantiating data are unavailable. The monthly variation depends to a large extent on certain physical characteristics of the reservoir. It might be expected that the monthly variation in the pan coefficient applicable to small shallow ponds would be small because the change in energy storage in such ponds, when expressed on a unit-area-time basis, is small compared with the net radiation received. For the oceans, the change in energy storage is a large item, and the resulting seasonal variation in evaporation lags considerably behind that of a pan. *Jacobs* [1943] computed that evaporation from both the North Pacific and North Atlantic Oceans is greatest in winter and least in summer, which is exactly opposite to the seasonal variation in pan evaporation. Thus, it is concluded that Mansfield's method of evaluating the success of his monomolecular film in suppressing evaporation may be satisfactory for small ponds, but it may be subject to considerable error if used for large reservoirs. Moreover, this method provides no reliable means of estimating the rise in water-surface temperature caused by the application of the film.

A better method of assessing the suppression of evaporation may be obtained by using a combination of the energy-budget and mass-transfer evaporation measurement techniques. The energy-budget equation is as follows

$$[Q_s - Q_r + Q_a - Q_{ar} + Q_i] - [Q_{bs} + Q_e + Q_h + Q_w] = Q_\delta \quad (1)$$

Inflow - Outflow = Change in storage  
in which

$Q_s$  = solar radiation incident to the water surface

$Q_r$  = reflected solar radiation

$Q_a$  = incoming long-wave radiation from the atmosphere

$Q_{ar}$  = reflected long-wave radiation

$Q_i$  = net energy advected into the body of water

$Q_{bs}$  = long-wave radiation emitted by the body of water

$Q_e$  = energy utilized by evaporation

$Q_h$  = energy conducted from the body of water as sensible heat

$Q_w$  = energy advected by the evaporated water

$Q_\delta$  = increase in energy stored in the body of water

Because of their small magnitude, certain items in the energy budget have been neglected. These include heating caused by chemical and biological processes, conduction of heat through the lake bottom, and transformation of kinetic into thermal energy. *Koberg* [1958], in the report on the Lake Mead studies, has described in detail how each term in the energy budget is evaluated.

It is assumed that the application of a monomolecular film will not affect the items grouped together as inflow in the energy-budget equation. It is evident that the film could have no effect whatever on  $Q_s$ ,  $Q_a$ , and  $Q_r$ . Optical tests indicate that the film has little effect on the reflectivity of the water surface. Any effect of a film on  $Q_{ar}$  would be counterbalanced to a large extent by a compensating change in  $Q_{bs}$ , as will be shown.

In connection with the Lake Hefner studies the emissivity of water was determined by Gier and Dunkle [Anderson, 1954] to be 0.970, and the back radiation term,  $Q_{bs}$ , was computed as follows

$$Q_{bs} = 0.970\sigma(T_0 + 273)^4 \quad (2)$$

in which

$\sigma$  = the Stefan-Boltzman constant for black-body radiation

$T_0$  = water-surface temperature in  $^{\circ}\text{C}$

The emissivity was found to be independent of water temperature and dissolved solids. Gier and Dunkle measured it indirectly. Instead of measuring the energy emitted by the water surface, they measured the reflectivity of the water surface using calibrated energy sources of various temperatures. According to Kirchhoff's law, the emissivity is precisely equal to the absorptivity when the source temperature is equal to the water-surface temperature. The absorptivity is one minus the reflectivity, thus the technique permits a measurement of the emissivity.

In the discussion to follow, the relative magnitude of the items in (1) are pertinent to the reasonableness of certain approximations and assumptions that will be made. The average annual values of the various items measured at Lake Hefner, Oklahoma, in 1950-1951 and at Lake Mead, Arizona-Nevada, in 1952-1953 as presented by *Koberg* [1958] are given in Table 1.

TABLE 1—Average daily values in cal cm<sup>-2</sup> of terms in energy-budget equation for Lake Mead and Lake Hefner

Reservoir	$Q_s$	$Q_r$	$Q_a$	$Q_{ar}$	$Q_{bs}$	$Q_v$	$Q_e$	$Q_h$	$Q_w$	$Q_g$
Lake Mead	506	37	692	20	842	-4	344	-5	12	-56
Lake Hefner	420	26	638	19	781	2	222	8	6	-2

A decrease of one per cent in emissivity does not affect  $Q_a$ , but increases  $Q_{ar}$  by one per cent of  $Q_a$ . It also decreases  $Q_{bs}$  by one per cent. In (1),  $Q_{ar}$  and  $Q_{bs}$  have the same signs, and as  $Q_a$  and  $Q_{bs}$  are not too greatly different in magnitude, the net effect of a one per cent error in emissivity is negligible compared with  $Q_e$ . However, once it has been established just which of the evaporation-suppressing alcohols is to be preferred, and also which solvent, if any, should be used, laboratory tests will be made to determine whether the film has any effect on emissivity.

It appears that the film has relatively little effect on the inflow items of (1). Over a long period of time, the film could have no appreciable effect on  $Q_g$ . The immediate effect of a film is to increase the water-surface temperature and, therefore, the amount of stored energy. Once the initial rise in water-surface temperature has taken place, however, the effect of a film on the change in stored energy is negligible. The amount of energy stored in the lake will be greater than if there were no film, but subsequent seasonal changes in energy storage will not be affected.

After equilibrium has been reached, the effect of the film on some of the outflow items of (1) will be appreciable. From Table 1,  $Q_w$  is a small item, particularly in comparison with  $Q_{bs}$ . If evaporation is reduced 25 pct by the use of a monomolecular film (which would be very desirable from a practical viewpoint) the resulting change in  $Q_w$  would be of little consequence.

The film has a marked effect on the three remaining items,  $Q_{bs}$ ,  $Q_e$ , and  $Q_h$ . As these three items are considered to be the only ones on which the film has an appreciable effect, it follows that the net sum of the effects must be zero, or

$$(Q_{bs}' - Q_{bs}) + (Q_e' - Q_e) + (Q_h' - Q_h) = 0 \quad (3)$$

in which the symbols with primes refer to the

reservoir with a film and the symbols without primes to the same reservoir without a film. Eq. (3) indicates that if evaporation is reduced ( $Q_e' - Q_e < 0$ ), there must be a corresponding increase in the sum  $(Q_{bs}' - Q_{bs}) + (Q_h' - Q_h)$ . If evaporation is decreased, the water-surface temperature must rise until the energy no longer being utilized for evaporation is returned to the atmosphere by increased back radiation and conduction.

Prior to the application of a film  $Q_h$  can be computed using the Bowen [1926] ratio, which is defined as the ratio  $Q_h/Q_e$ . Differences of opinion have persisted for many years concerning the validity of the Bowen ratio. The Lake Hefner studies [Anderson, 1954] and the Lake Mead studies [Koberg, 1958] both indicated that regardless of any theoretical shortcomings, the Bowen ratio provided a satisfactory means of determining  $Q_h$ , at least for the purpose of computing evaporation under ordinary meteorological conditions. The Bowen ratio is

$$R = c \frac{(T_0 - T_a)P}{(e_0 - e_a)1000} \quad (4)$$

in which

$R$  = Bowen ratio

$c$  = a coefficient having a most probable value of 0.61, according to Bowen

$T_a$  = air temperature, °C

$e_0$  = saturation vapor pressure at the temperature of the water surface, millibars

$e_a$  = vapor pressure of the air, millibars

Koberg [1958] showed that for the purpose of computing evaporation, the height at which both  $T_a$  and  $e_a$  were measured was immaterial. Thus, before the film is applied, evaporation may be computed using (1) for period between thermal surveys, and for each of those periods the average value of  $Q_h$  may be determined.

After the film has been applied, the Bowen ratio is no longer valid. The film is applied expressly to reduce  $Q_e$ , and as previously mentioned, both  $Q_h$  and  $Q_{bs}$  must then increase. Then  $Q_h$  must be determined from the heat-transfer equation, which is

$$Q_h = c_p A (dT/dz + \gamma) \quad (5)$$

in which

$c_p$  = specific heat of air at constant pressure



$A$  = Austausch, or exchange coefficient

$T$  = temperature in  $^{\circ}\text{C}$

$z$  = height in cm

$\gamma$  = adiabatic lapse rate,  $0.98^{\circ}\text{C}$  per 100 meters

As  $\gamma$  is generally small compared with  $dT/dz$ , it may be neglected. The derivative  $dT/dz$  may be replaced by finite differences. The exchange coefficient  $A$  need not be precisely defined; it may be assumed to be proportional to the wind speed  $u$ . Eq. (5) then becomes

$$Q_h = Ku(T_0 - T_a) \quad (6)$$

in which  $K$  is an empirical constant containing among other things  $c_p$ ,  $z$ , the height at which  $T_a$  and  $u$  are measured, and any variables other than  $u$  contained in  $A$ .

During the pretreatment or calibration period, it is necessary to obtain records of  $T_0$  and  $T_a$  to compute certain items in the energy-budget equation. If a record of  $u$  is also obtained, all the data are available that are needed to evaluate  $K$  in (6). After the film has been applied,  $Q_h'$  can be computed using this equation.

During the calibration period  $Q_e$  is to be determined using (1). From data obtained at that time, the value of  $N$  in the mass-transfer equation

$$Q_e = Nu(e_0 - e_a) \quad (7)$$

is to be determined.

After the film has been applied to the lake,  $Q_e'$  can be computed only from (1). Eq. (7) cannot be used because the film prevents the unimpeded movement of water molecules away from the water surface.  $Q_h'$  must be determined from (6); the Bowen ratio cannot be used. It then remains to determine what the evaporation would have been had no film been applied. Combining (2), (6), and (7) with (3) we obtain

$$\begin{aligned} [Q_{b_s}' - 0.970\sigma(T_0 - 273)^4] \\ + [Q_e' - Nu(e_0 - e_a)] \\ + [Q_h' - Ku(T_0 - T_a)] = 0 \end{aligned} \quad (8)$$

$Q_{b_s}'$ ,  $Q_e'$ , and  $Q_h'$  have been computed. Eq. (8), which contains only one unknown,  $T_0$ , was obtained by combining heat and mass transfer equations with the energy-budget equation. The saturation vapor pressure  $e_0$  is explicitly determined by  $T_0$ , but the mathematical relationship is not simple. A trial-and-error solution is used

with an assumed value of  $T_0$  and the corresponding tabular value of  $e_0$ . When a value of  $T_0$  is found that will cause (8) to balance,  $Q_e$  can be computed using (7) and compared with  $Q_e'$  to determine the reduction in evaporation.

This method was used to compute the effect on evaporation resulting from the application of a film to a small ranch pond near San Antonio, Texas. Octadecanol in solid form was applied by personnel of Southwest Research Institute; R. R. Cruse of the Institute also maintained the radiation equipment and obtained the necessary field data. During the period October 10-31, 1956, the computed reduction in evaporation was 18 pct. The observed water-surface temperature was  $24.3^{\circ}\text{C}$ . The water-surface temperature would have been  $22.4^{\circ}\text{C}$  had no film been applied; therefore the 18 pct reduction in evaporation was necessarily accompanied by a rise of  $1.9^{\circ}\text{C}$ .

During the calibration period and also after the film is applied, the basic instrumentation required includes a total hemispherical radiometer, a pyrheliometer, a psychrometer, a water-surface temperature recorder, an underwater thermometer, and an anemometer. Accessories such as boats, rafts, and suitable recorders are also needed.

It may appear that instrumental requirements for the above-described method are unduly complicated. The equipment is expensive compared with an evaporation pan, but techniques have been developed to record nearly all of the data on one multichannel recording potentiometer. All equipment is designed to operate without attention between weekly servicing visits.

Basically the method involves the computation of two unknowns: the evaporation and the water-surface temperature that would have been observed had no film been applied. To permit solving for two unknowns, two simultaneous equations are provided, one based on the energy-budget concept, and the other upon mass-transfer theory. A basic heat-transfer equation is used instead of the Bowen ratio to determine the energy conducted to the atmosphere as sensible heat after the film has been applied.

## REFERENCES

- ANDERSON, E. R., Energy-budget studies, *water-loss investigations: Lake Hefner studies, tech. rep.*, U. S. Geol. Surv. Prof. Pap. 269, pp. 71-119, 1954.

- BOWEN, I. S., The ratio of heat losses by conduction and by evaporation from any water surface, *Phys. Rev.*, **27**, 779-787, 1926.
- JACOBS, W. C., Sources of atmospheric heat and moisture over the North Pacific and North Atlantic Oceans, *Annals N. Y. Acad. Sci.*, **44**, 19-40, 1943.
- KOBERG, G. E., Energy-budget studies, *water-loss investigations, Lake Mead studies*, U. S. Geol. Surv. Prof. Pap. 298, pp. 20-29, 1958.
- KOHLER, M. A., Lake and pan evaporation, *water-loss investigations, Lake Hefner studies, tech. rep.*, U. S. Geol. Surv. Prof. Pap. 269, pp. 127-148, 1954.
- LANGMUIR, I., AND D. B. LANGMUIR, The effect of monomolecular films on the evaporation of ether solutions, *J. Phys. Chem.*, **31**, 1719-1731, 1927.
- LANGMUIR, I., AND V. J. SCHAEFER, Rates of evaporation of water through compressed monolayers on water, *J. Franklin Inst.*, **235**, 119-162, 1943.
- MANSFIELD, W. W., Effect of surface films on the evaporation of water, *Nature* **172**, no. 4389, 1101, 1953.
- MANSFIELD, W. W., The use of hexadecanol for reservoir evaporation control, *Proc. 1st Internat. Conf. Reservoir Evap. Control*, Southwest Research Institute, pp. 33-35, 1956.
- RIDEAL, E. K., The influence of thin surface films on the evaporation of water, *J. Phys. Chem.*, **29**, 1585-1588, 1925.

(Manuscript received July 9, 1958.)



# A Note on the Propagation of a Sound Pulse in a Two-Layer Liquid Medium\*

J. H. ROSENBAUM

Shell Development Company  
Houston, Texas

**Abstract**—The contribution from the branch line integral in Pekeris' formulation of the propagation of explosive sound in a two-layer liquid medium is investigated. It is shown that there is no *net* contribution to the solution in terms of pressure amplitude measured in the upper layer, which decays less rapidly than  $r^{-2}$ . The horizontal range, designated by  $r$ , is large compared with the depth of the layer. This result is applicable to early arrival pressure signals and is in agreement with recently published model experiments.

**Introduction**—In a recent article, *Knudsen* [1957] has investigated experimentally the propagation of a pressure transient in a liquid layer overlying an infinitely deep liquid bottom. He finds that the pressure amplitude of a first-arrival peak at large horizontal distances,  $r$ , decays as  $r^{-2}$ . This is in agreement with an expression for the "branch line integral" given by *Pekeris* [1948], but does not agree with an  $r^{-1}$  contribution for the integral, derived more recently by *Officer* [1953]. The leading term of the "normal mode" contribution, derived by *Pekeris*, decays as  $r^{-1}$ , but arrives with vanishing amplitude. (The Airy phase, which represents a late signal, decays as  $r^{-5/6}$ .) It is the purpose of this note to point out that *Knudsen's* experimental results are indeed in agreement with theory, and that such a first arrival at large horizontal ranges cannot decay less rapidly than  $r^{-2}$ .

To facilitate comparison, we retain, except for some minor changes, the notation and paths of integration used by *Pekeris* and *Officer*. The reader is referred to these papers for more details.

We introduce cylindrical coordinates  $r, z$  with origin directly above the source at the free surface of the layer and the positive  $z$  axis pointing downward into the fluid medium.  $r$  is the radial coordinate parallel to the surface; because of symmetry, the problem is independent of the polar angle.

We let  $c_2$  and  $c_1$  be the sound velocities respectively in the liquid half space and the liquid layer of depth  $H$ , with  $c_2 > c_1$ , which is the case of interest in refraction shooting.  $b$  is the ratio of

the density of the layer to that of the half space;  $\omega$  is the angular frequency;  $k$  is the wave number in the radial direction; and  $t$  is the time.

An acoustic pressure point source at  $r = 0$  and  $z = d < H$ , emits a pressure pulse at  $t = 0$ , described by a peak amplitude factor (at unit distance)  $A$  and an exponential decay time constant  $\theta$ . The pressure signal  $P$ , received in the layer at a point  $(r, z)$  may then be expressed as the sum of two double integrals; one,  $P_N$ , is taken around the "normal mode" poles on the real  $k$  axis ( $\omega/c_2 < k < \omega/c_1$ ), and the other,  $P_B$ , is taken around a branch line  $\Gamma$ , as shown in Figure 1. We ignore here the contributions from poles in the shaded region of the figure, because they lead to expressions which decay exponentially with  $r$ .

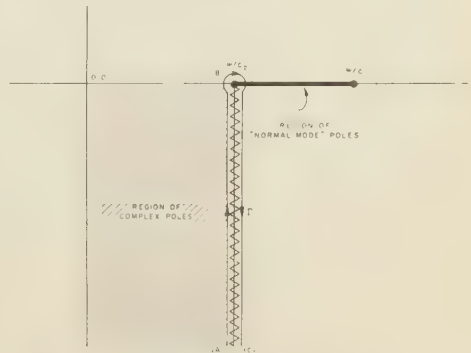


FIG. 1—The top sheet of the  $k$  plane. The original path of integration proceeds from  $-i \infty$  to 0 to  $+\infty$  above the branch point and poles on the real axis. This path is distorted into a quarter circle with infinite radius in the fourth quadrant, whose contribution to the integral vanishes.

\* Publication No. 169, Shell Development Company, Exploration and Production Research Division, Houston, Texas.



The branch line integral—We may write the branch line integral\* in the form

$$P_B = \frac{A}{\pi} \operatorname{Re} \int_0^\infty \frac{e^{i\omega t}}{(1/\theta + i\omega)} d\omega \\ \cdot \int_\Gamma H_0^{(2)}(kr) \frac{\sin \beta_1 d}{\beta_1} F_1(\omega, k) dk \quad (1)$$

where

$$F_1(\omega, k) \\ = \frac{\beta_1 \cos \beta_1(H - z) + ib\beta_2 \sin \beta_1(H - z)}{\beta_1 \cos \beta_1 H + ib\beta_2 \sin \beta_1 H}$$

for  $(d \leq z \leq H)$ .\*\*

$$\beta_1 = \left( \frac{\omega^2}{c_1^2} - k^2 \right)^{\frac{1}{2}} = -i \left( k^2 - \frac{\omega^2}{c_1^2} \right)^{\frac{1}{2}} = -i\tilde{\beta}_1;$$

$$\beta_2 = \left( \frac{\omega^2}{c_2^2} - k^2 \right)^{\frac{1}{2}} = -i \left( k^2 - \frac{\omega^2}{c_2^2} \right)^{\frac{1}{2}} = -i\tilde{\beta}_2.$$

$\tilde{\beta}_1$  and  $\tilde{\beta}_2$  are positive and real on the real  $k$  axis for  $k > (\omega/c_1)$  and  $k > (\omega/c_2)$  respectively; they are positive and imaginary for  $k < (\omega/c_1)$  and  $k < (\omega/c_2)$ . This defines the top sheet of the  $k$  plane as illustrated in Figure 1. For the Hankel function  $H_0^{(2)}(kr)$ , we substitute the first term of its asymptotic expansion, valid for large values of the argument  $kr$ , introduce a change in variable  $q = ik - i\omega/c_2$ , and combine the two parts of the integral along  $\Gamma$  ( $AB$  and  $BC$ ) into one (along  $BC$ ). We obtain, on changing the order of integration and some algebraic reduction,

$$P_B = -\frac{2A}{\pi} \operatorname{Im} \int_0^\infty e^{-r q} dq \\ \cdot \int_0^\infty \frac{e^{i\omega(t-r/c_2) + i\pi/4}}{(1/\theta + i\omega)} \left( \frac{2k}{\pi r} \right)^{\frac{1}{2}} F_2(\omega, q) d\omega \quad (2)$$

where

$$F_2(\omega, q) = \frac{b\tilde{\beta}_2 \sin \beta_1 z \sin \beta_1 d}{(\beta_1 \cos \beta_1 H + b\tilde{\beta}_2 \sin \beta_1 H)} \\ \cdot \frac{1}{(\beta_1 \cos \beta_1 H - b\tilde{\beta}_2 \sin \beta_1 H)}.$$

The phase of  $\tilde{\beta}_2$  is that along the path ( $BC$ ) and varies from  $-\pi/4$  to  $-\pi/2$  as  $q$  increases

from 0 to  $\infty$ . When  $q = 0$ ,  $\tilde{\beta}_2$  and therefore the numerator of  $F_2(\omega, q)$  vanish, and both factors in the denominator of  $F_2(\omega, q)$  become equal to  $\beta_1^0 \cos \beta_1^0 H$ , where the superscript 0 refers to an evaluation of the function at the point  $q = 0$ . The denominator of  $F_2(\omega, 0)$  vanishes when  $\cos \beta_1^0 H = 0$ , and the integrand over  $\omega$  in (2) goes to infinity at those points. When  $q > 0$ , neither factor in the denominator of  $F_2(\omega, q)$  can vanish for real values of  $\omega$ .\* Since the integrals over  $q$  and  $\omega$  then converge, the change of order of integration, which was carried out above, is permissible if we define the integration over  $q$  to extend from  $\delta > 0$  to  $\infty$  and let  $\delta$  approach zero in the limit.

The definition of  $\tilde{\beta}_2$ , as chosen above, is  $\tilde{\beta}_2 = q^{\frac{1}{2}}(\omega - i\frac{1}{2}c_2q)^{\frac{1}{2}}(2/c_2)^{\frac{1}{2}}e^{-i\pi/4}$ . For real  $q > 0$ ,  $\tilde{\beta}_2$  has a branch point at  $+i\frac{1}{2}c_2q$  in the complex  $\omega$  plane, and  $k^{\frac{1}{2}}$  has one at  $+ic_2q$ . We cut the  $\omega$  plane between  $i\frac{1}{2}c_2q$  and  $+i\infty$  along the positive imaginary  $\omega$  axis. It is easily shown that of the two factors in the denominator of  $F_2(\omega, q)$ , the first one has roots in the second and fourth quadrants of the complex  $\omega$  plane, while the second one has roots in the first and third quadrants. Since  $t > (r/c_2)$ ,\*\* the path of integration over  $\omega$  in (2) can be shifted from the real  $\omega$  axis to the positive imaginary axis to the right of the branch cut and the pole at  $i/\theta$ . To this integral, which extends from  $\omega = 0$  to  $\omega = +i\infty$ , we must of course add the contributions from the poles of  $F_2(\omega, q)$  which are in the first quadrant. We obtain consequently

$$P_B = -\frac{2A}{\pi} \operatorname{Im} \int_0^\infty e^{-r q} dq \\ \cdot \int_0^{+i\infty} \frac{e^{i\omega(t-r/c_2) + i\pi/4}}{(1/\theta + i\omega)} \left( \frac{2k}{\pi r} \right)^{\frac{1}{2}} F_2(\omega, q) d\omega \\ - \frac{2A}{\pi} \operatorname{Im} \int_0^\infty e^{-r q} dq \\ \cdot 2\pi i \sum_n \frac{e^{i\omega_n(t-r/c_2) + i\pi/4}}{(1/\theta + i\omega_n)} \left( \frac{2k_n}{\pi r} \right)^{\frac{1}{2}}$$

\* This is true because, as stated in the introduction, we ignore complex poles in the fourth quadrant of the  $k$  plane. Otherwise, it appears possible for poles to occur on the branch line,  $q$  real and  $> 0$  ( $\omega$  real positive). See Appendix.

\*\*  $r$  is so much greater than  $H$  that  $t = r/c_2$  may be considered to be the time of the first refracted arrival. We note here that all expressions in this report apply to  $t > r/c_2$  and can be shown to vanish for  $t < r/c_2$ .

\* See equations (A23) and (A48) of Pekeris [1948], or equations (26) and (35) of Officer [1953].

\*\* Interchange  $z$  and  $d$  for  $(0 \leq z \leq d)$ .

$$\begin{aligned}
 & \frac{b\bar{\beta}_{2n} \sin \beta_{1n} z \sin \beta_{1n} d}{(\beta_{1n} \cos \beta_{1n} H + b\bar{\beta}_{2n} \sin \beta_{1n} H)} \\
 & \cdot \frac{1}{[(\partial/\partial\omega)(\beta_{1n} \cos \beta_{1n} H - b\bar{\beta}_{2n} \sin \beta_{1n} H)]_n} \\
 & = P_{B_1} + P_{B_2}. \quad (3)
 \end{aligned}$$

The summation extends over  $\omega = \omega_n$  and refers to the solutions ( $n = 1, 2, \dots$ ) of the equation  $f_2 = \beta_{1n} \cos \beta_{1n} H - b\bar{\beta}_{2n} \sin \beta_{1n} H = 0$ , which provides a relation between  $\omega_n$  and  $q$ . The subscript  $n$  also refers to an evaluation of a function at  $\omega = \omega_n$ . The partial derivative,  $\partial f_2/\partial\omega$ , in the denominator of the second term,  $P_{B_2}$ , is to be taken at constant  $q$ .

First we evaluate the contribution of the double integral  $P_{B_1}$ . We set  $\omega = i\xi$  and change the order of integration. This is permissible since both integrals are convergent, and the pole at  $\xi = 1/\theta$  is avoided by means of an indentation in the path of integration.

$$\begin{aligned}
 P_{B_1} = & -\frac{2A}{\pi} \operatorname{Re} \int_0^\infty \frac{e^{-\xi(l-r/c_2) + i\pi/4}}{(1/\theta - \xi)} d\xi \\
 & \cdot \int_0^\infty e^{-r\alpha} \left(\frac{2k}{\pi r}\right)^{\frac{1}{2}} F_2(\xi, q) dq \quad (4)
 \end{aligned}$$

For large values of  $r$ , only small values of  $q$  contribute. Expansion in powers of  $q$  leads to an asymptotic solution of the integral in inverse powers of  $r$ , whose leading term yields after some cancelling

$$\begin{aligned}
 P_{B_1} = & -\frac{2Ab}{\pi r^2} \operatorname{Im} \int_0^\infty \frac{e^{-\xi(l-r/c_2)}}{(\xi - 1/\theta)} \\
 & \cdot \frac{\sinh \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{\xi z}{c_2} \right] \sinh \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{\xi d}{c_2} \right]}{\frac{\xi}{c_2} \left( \frac{c_2^2}{c_1^2} - 1 \right) \cosh^2 \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{\xi H}{c_2} \right]} d\xi \quad (5)
 \end{aligned}$$

Now the integrand of this expression is purely real, except for the indentation around the pole at  $\xi = 1/\theta$ . Evaluation of this contribution finally yields

$$\begin{aligned}
 P_{B_1} = & -\frac{2Ab}{r^2} e^{-(l-r/c_2)/\theta} \\
 & \cdot \frac{\sinh \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{z}{c_2 \theta} \right] \sinh \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{d}{c_2 \theta} \right]}{\left( \frac{c_2^2}{c_1^2} - 1 \right) \cosh^2 \left[ \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}} \frac{H}{c_2 \theta} \right]} c_2 \theta \quad (6)
 \end{aligned}$$

which is seen to decay as  $r^{-2}$ .

Next we evaluate the contribution of the integral  $P_{B_2}$ . We change the integration over  $q$  to one over  $\omega_n$ . The relation between  $q$  and  $\omega_n$  is given by the equation  $f_2 = 0$ . We see that

$$\begin{aligned}
 dq = & \left( \frac{dq}{d\omega} \right)_{f_2} d\omega = -\frac{(\partial f_2/\partial\omega)_q}{(\partial f_2/\partial q)_\omega} d\omega \\
 = & -i \frac{(\partial f_2/\partial\omega)_q}{(\partial f_2/\partial k)_\omega} d\omega \quad (7)
 \end{aligned}$$

and consequently  $P_{B_2}$  becomes

$$\begin{aligned}
 & -2A \sum_n \operatorname{Im} \int_{\omega_n^0}^{i\infty + c_2 k_n} \frac{e^{i(\omega_n t - k_n r) + i\pi/4}}{(1/\theta + i\omega_n)} \left( \frac{2k_n}{\pi r} \right)^{\frac{1}{2}} \\
 & \cdot \frac{\sin \beta_{1n} z \sin \beta_{1n} d}{\sin \beta_{1n} H [\partial/\partial k (\beta_{1n} \cos \beta_{1n} H - b\bar{\beta}_{2n} \sin \beta_{1n} H)]_n} d\omega_n. \quad (8)
 \end{aligned}$$

The change of order of integration and summation is most readily justified by the physical consideration that the number of modes to be taken into account in practice is finite. The partial derivative,  $\partial f_2/\partial k$ , in (8) refers to constant  $\omega$ . The lower limit, corresponding to  $q \rightarrow 0$ , is given by

$$\omega \rightarrow \omega_n^0 = c_2 k_n^0 = \frac{(n - \frac{1}{2})c_2 \pi}{H \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}}}.$$

We also note that as  $q \rightarrow 0$ ,

$$\left( \frac{d\omega}{dq} \right)_{f_2} \rightarrow e^{i3\pi/4} \frac{c_2^{\frac{3}{2}} b}{H \sqrt{2\omega_n^0 q} \left( \frac{c_2^2}{c_1^2} - 1 \right)}.$$

It is evident, especially from the earlier expressions in terms of  $q$  (which we could have integrated directly), that the main contribution to the integral must arise from the neighborhood of the lower limit of integration. We obtain this contribution by expanding around the end-point of the interval of integration. We follow the method described by *van der Corput* [1936] and write

$$\begin{aligned}
 P_{B_2} = & -2A \sum_n \operatorname{Im} \int_0^{\omega_n^0 + i\infty} \\
 & \left[ Q_1(\omega_n^0) + Q_1'(\omega_n^0)h + Q_1''(\omega_n^0) \frac{h^2}{2!} + \dots \right] \\
 & \cdot \left[ 1 + Q_2'''(\omega_n^0) \frac{h^3}{3!} + \dots \right] \exp [Q_2(\omega_n^0) \\
 & + Q_2'(\omega_n^0)h + Q_2''(\omega_n^0)(h^2/2)] dh \quad (9)
 \end{aligned}$$

where

$$Q_1(\omega_n) = \frac{e^{i\pi/4}}{(1/\theta + i\omega_n)} \left( \frac{2k_n}{\pi r} \right)^{\frac{1}{2}} \cdot \frac{\sin \beta_{1n} z \sin \beta_{1n} d}{\sin \beta_{1n} H [\partial/\partial k (\beta_1 \cos \beta_1 H - b\bar{\beta}_2 \sin \beta_1 H)]_n},$$

$$Q_2(\omega_n) = i(\omega_n t - k_n r),$$

and the primes denote total differentiation with respect to  $\omega_n$ , that is, under the condition that  $f_2 = 0$ . We evaluate at the lower limit,  $\omega_n = \omega_n^0$ ,

$$Q_1(\omega_n^0) = 0$$

$$Q_1'(\omega_n^0) = \frac{e^{i\pi/4}}{(1/\theta + i\omega_n^0)} \left( \frac{2k_n^0}{\pi r} \right)^{\frac{1}{2}} \cdot \frac{\sin \beta_{1n}^0 z \sin \beta_{1n}^0 d}{b^{\frac{2}{3}} c_2} H \left( \frac{c_2^2}{c_1^2} - 1 \right) \quad \text{and}$$

$$Q_2(\omega_n^0) = i\omega_n^0(t - r/c_2)$$

$$Q_2'(\omega_n^0) = i(t - r/c_2)$$

$$Q_2''(\omega_n^0) = -i \frac{\omega_n^0 H^2}{c_2^2 b^{\frac{2}{3}}} \left( \frac{c_2^2}{c_1^2} - 1 \right)^2 r.$$

The integral (9) is now expressible in terms of Fresnel's integrals. For  $r/c_2 \gg (t - r/c_2)$ , we obtain as the leading term of an asymptotic series in negative powers of  $r$ ,

$$P_{B_2} = \frac{1}{r^{\frac{1}{2}}} - \frac{2\sqrt{2}c_2^{\frac{1}{2}}}{\sqrt{\pi} H \left( \frac{c_2^2}{c_1^2} - 1 \right)} \cdot \sum_n \frac{\sin \beta_{1n}^0 z \sin \beta_{1n}^0 d}{(\omega_n^0)^{\frac{1}{2}} [1/\theta^2 + (\omega_n^0)^2]^{\frac{1}{2}}} \cdot \cos [\omega_n^0(t - r/c_2) + \epsilon], \quad (10)$$

where

$$\omega_n^0 = \left( n - \frac{1}{2} \right) c_2 \pi / H \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}},$$

$$\beta_{1n}^0 = (n - \frac{1}{2})\pi/H,$$

and

$$\epsilon = [\pi/4 - \tan^{-1}(\omega_n^0 \theta)].$$

Contributions from any other critical points, as defined by van der Corput, in the interval of integration of (8) would decay exponentially with distance  $r$  and may be neglected here.

If we regard the branch line integral  $P_B$  in

the original  $k$  plane, with  $\omega$  real and positive, then we see that the expressions (9) and (10) for  $P_B$  represent contributions which are due to the fact that there are poles in the neighborhood of the branch point. The poles in question are the "normal mode" poles (see Figure 1), whose behavior near the branch point is readily determined. We note that the period equation  $f = \beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H = 0$ , which defines the poles, has continuous solutions for real values of  $k$  and  $\omega$  on both Riemann sheets of the  $k$  plane. At  $q = 0$ ,

$$(dk/d\omega)_{q=0} = c_2^{-1} \quad \text{and} \quad (d^2k/d\omega^2)_{q=0} = 0,$$

while

$$(dk/d\omega)_f = c_2^{-1}$$

$$(d^2k/d\omega^2)_f = \frac{\omega H^2}{b^{\frac{2}{3}} c_2^{\frac{2}{3}}} \left( \frac{c_2^2}{c_1^2} - 1 \right)^2,$$

where the subscript  $f$  refers to a solution of the period equation. Thus with increasing  $k$ , the normal mode poles pass through the branch point  $k = \omega/c_2$  to the upper  $\beta_2$  sheet. We shall discuss these poles again in the next section.

*The end-point contribution*—The expression (10) for  $P_{B_2}$ , which also represents the leading term of an asymptotic expansion for the branch line integral contribution,  $P_B$ , to the pressure signal of the present problem, does not represent the entire "early arrival" pressure signal received in the layer at large horizontal ranges  $r$ , [ $r \gg H$ ], and early times  $t$ , [ $0 < (t - r/c_2) \ll r/c_2$ ]. To it we must add a pressure signal,  $P_{N_1}$ , which is due to the "normal mode" poles on the real  $k$  axis, as indicated in Figure 1. This normal mode pole contribution has been discussed by Pekeris [1948] and others and is given by the equation\*

$$P_N = +2A \operatorname{Im} \int_0^\infty \frac{e^{i\omega t}}{(1/\theta + i\omega)} \quad (11)$$

$$\cdot \sum_n \operatorname{Res.} \left[ H_0^{(2)}(kr) \frac{\sin \beta_1 d}{\beta_1} F_1(\omega, k) k \right] d\omega$$

where  $\operatorname{Res.}$  designates the residue at the normal mode poles, and the summation refers to the solutions ( $n = 1, 2, \dots$ ) of the period equation,  $f = \beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H = 0$ , in the denominator of  $F_1(\omega, k)$ . As was pointed out at the end

\* See equations (A23) and (A48) of Pekeris [1948].

of the last section, these zeros arise on the top sheet of the  $k$  plane only for  $k > \omega/c_2$ , so that the integration over  $\omega$  in (11) actually extends from

$$\omega = \omega_n^0 = c_2 k_n^0 = \left( n - \frac{1}{2} \right) c_2 \pi / H \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}}$$

to infinity.

Introducing the first term of the asymptotic expansion of the Hankel function, we can, after some reduction, write (11) as

$$P_N = +2A \sum_n Im \int_{\omega_n^0}^{\infty} \frac{e^{i(\omega_n t - k_n r) + i\pi/4}}{(1/\theta + i\omega_n)} \left( \frac{2k_n}{\pi r} \right)^{\frac{1}{2}} \cdot \frac{\sin \beta_{1n} z \sin \beta_{1n} d}{\sin \beta_{1n} H [\partial/\partial k (\beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H)]_n} d\omega_n, \quad (12)$$

where  $(\partial/\partial k)$  refers to constant  $\omega$ , and the subscripts  $n$  indicate evaluation at the roots of the period equation,  $f = 0$ . Pekeris evaluates this integral asymptotically for large values of  $r$  by the method of stationary phase. The leading term of his expansion decays as  $r^{-1}$ , except for the last arrival "Airy phase," which decays as  $r^{-5/6}$ . As  $(t - r/c_2) \rightarrow 0$ , that is, for very early signals, the leading term of his solution vanishes because of a factor  $\bar{\beta}_2$  in the numerator. This indicates, of course, a leading term of higher order than  $r^{-1}$  as  $(t - r/c_2) \rightarrow 0$ , which we must now evaluate.

In the interval of integration from  $\omega_n^0$  to  $\infty$ , the following points are of importance for the evaluation of the integral in asymptotic form for large values of  $r$ : the end point at  $\omega_n^0$ ; the point of stationary phase, where the first derivative of the exponent of  $e$  vanishes; and a stationary point of order two, corresponding to the "Airy phase" discussed by Pekeris, where the second derivative of the exponent of  $e$  also vanishes. The end-point at  $\omega = \infty$  cannot contribute to the integral, and since we are interested only in early arrival signals, we may ignore the stationary point of order two, that is, the late arrival Airy phase. As  $(t - r/c_2) \rightarrow 0$ ,  $(dk_n/d\omega_n)$ , which represents the reciprocal of the group velocity, approaches  $1/c_2$ , and the point of stationary phase approaches the end-point,  $\omega_n^0$ . To obtain the leading terms in the present approximation, we need consider only the end-point contribution at  $\omega_n^0$ , and following van der Corput, we write

$$P_N = +2A \sum_n Im \int_0^{\infty} e^{-i\pi/4} \cdot \left[ Q_3(\omega_n^0) + Q_3'(\omega_n^0)h + Q_3''(\omega_n^0) \frac{h^2}{2!} + \dots \right] \cdot \left[ 1 - Q_2'''(\omega_n^0) \frac{h^3}{3!} + \dots \right] \exp [Q_2(\omega_n^0) + Q_2'(\omega_n^0)h + Q_2''(\omega_n^0)(h^2/2)] dh, \quad (13)$$

where  $Q_2 = i(\omega_n t - k_n r)$ , and

$$Q_3 = \frac{e^{i\pi/4}}{(1/\theta + i\omega_n)} \left( \frac{2k_n}{\pi r} \right)^{\frac{1}{2}} \cdot \frac{\sin \beta_{1n} z \sin \beta_{1n} d}{\sin \beta_{1n} H [\partial/\partial k (\beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H)]_n}.$$

The primes denote total differentiation with respect to  $\omega_n$ , that is, with the condition that  $f = \beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H = 0$ . We evaluate at the lower limit,  $\omega_n = \omega_n^0$ ,

$$Q_3(\omega_n^0) = 0$$

$$Q_3'(\omega_n^0) = \frac{e^{i\pi/4}}{(1/\theta + i\omega_n^0)} \left( \frac{2k_n^0}{\pi r} \right)^{\frac{1}{2}} \cdot \frac{\sin \beta_{1n}^0 z \sin \beta_{1n}^0 d}{b^2 c_2} H \left( \frac{c_2^2}{c_1^2} - 1 \right)$$

$$Q_2(\omega_n^0) = i\omega_n^0(t - r/c_2)$$

$$Q_2'(\omega_n^0) = i(t - r/c_2)$$

$$Q_2''(\omega_n^0) = -i \frac{\omega_n^0 H^2}{c_2^3 b^2} \left( \frac{c_2^2}{c_1^2} - 1 \right)^2 r.$$

The integral (13) is again expressible in terms of Fresnel's integrals. For  $r/c_2 \gg (t - r/c_2)$ , we obtain as the leading term of an asymptotic series in negative powers of  $r$ ,

$$P_N = -\frac{A}{r^{\frac{1}{2}}} \frac{2\sqrt{2}c_2^{\frac{1}{2}}}{\sqrt{\pi} H \left( \frac{c_2^2}{c_1^2} - 1 \right)} \cdot \sum_n \frac{\sin \beta_{1n}^0 z \sin \beta_{1n}^0 d}{(\omega_n^0)^{\frac{1}{2}} [1/\theta^2 + (\omega_n^0)^2]^{\frac{1}{2}}} \cdot \cos [\omega_n^0(t - r/c_2) + \epsilon]. \quad (14)$$

This expression is exactly the negative of expression (10), which represents the leading term of  $P_B$ . Consequently, the leading terms of  $P_B$  and  $P_N$  cancel, and there is no net contribu-



tion to the first-arrival pressure signal which decays as  $r^{-\frac{1}{2}}$ .

*The first-arrival pressure signal*—Because of the cancellation of the leading terms of the branch line integral and normal mode end-point contributions, the first-arrival pressure signal is of the order  $r^{-2}$ . It is given by the expression (6) for  $P_{B_1}$ , and the second-order terms of expressions (9) and (13) for  $P_{B_2}$  and  $P_N$ . The latter two terms are identical and add. Their evaluation yields

$$\begin{aligned}
 P_{B_2} + P_N = & \frac{Ac_2}{r^2 \left( \frac{c_2^2}{c_1^2} - 1 \right)} \\
 & \cdot \sum_n \left\{ \sin \beta_{1n}^0 z \sin \beta_{1n}^0 d \right. \\
 & \left. \left[ \frac{4}{b} + \frac{3bc_2^2}{(\omega_n^0)^2 H^2 \left( \frac{c_2^2}{c_1^2} - 1 \right)} \right] \right. \\
 & \cdot \frac{\cos [\omega_n^0 (t - r/c_2) - \psi]}{[1/\theta^2 + (\omega_n^0)^2]^{\frac{1}{2}}} \\
 & \left. + \frac{4bc_2^2}{\omega_n^0 H^2 \left( \frac{c_2^2}{c_1^2} - 1 \right)} \frac{\cos [\omega_n^0 (t - r/c_2) + 2\epsilon]}{[1/\theta^2 + (\omega_n^0)^2]} \right\} \\
 & - \frac{4bc_2(z \cos \beta_{1n}^0 z \sin \beta_{1n}^0 d + d \cos \beta_{1n}^0 d \sin \beta_{1n}^0 z)}{\omega_n^0 H^2 \left( \frac{c_2^2}{c_1^2} - 1 \right)^{\frac{1}{2}}} \\
 & \cdot \frac{\cos [\omega_n^0 (t - r/c_2) - \psi]}{[1/\theta^2 + (\omega_n^0)^2]^{\frac{1}{2}}}, \quad (15)
 \end{aligned}$$

where  $\psi = \tan^{-1}(\omega_n^0 \theta)$ . Expressions (6) and (15) thus represent the leading terms of the first-arrival pressure signal, which decays with horizontal range as  $r^{-2}$ .<sup>\*</sup> Such a decay is, of course, the one we expect from other considerations, such as a solution of the problem by the ray theory expansion of Pekeris, which effectively reduces the layered problem to one involving two half spaces in contact at a plane.

The derivation of (14), as presented in the last section, demonstrates the importance of including

<sup>\*</sup> To obtain higher order terms than these, we must also consider higher order terms of the asymptotic expansion for  $H_0^{(2)}(kr)$ .

contributions due to end-points in the asymptotic evaluation of integrals by the method of stationary phase. However, a more direct solution to our problem can be obtained if we combine (8) and (12). The expression  $f_2 = \beta_1 \cos \beta_1 H - b\bar{\beta}_2 \sin \beta_1 H = 0$ , in the denominator of (8), is nothing but the period equation  $f = \beta_1 \cos \beta_1 H + b\bar{\beta}_2 \sin \beta_1 H = 0$ , on the lower Riemann surface. We need only to change the phase of  $\bar{\beta}_2$  in  $f_2$  by an angle  $\pi$  to make the expressions  $f$  and  $f_2$  identical. Indeed, if we had chosen the path (AB) instead of (BC) of Figure 1, in the representation (2) of the branch line integral, we would have arrived directly at such a form of equation (8). We can thus write, in somewhat more explicit form,

$$\begin{aligned}
 P_{B_2} + P_N \\
 = 2A \sum_n Im \int_{i\omega + c_2 k_n}^{\infty} \frac{e^{i(\omega_n t - k_n r)} + i\pi/4}{(1/\theta + i\omega_n)} \left( \frac{2k_n^{-1}}{\pi r} \right)^{\frac{1}{2}} \\
 \cdot \frac{\sin \beta_{1n} z \sin \beta_{1n} d (\beta_{1n}^2 + b^2 \bar{\beta}_{2n}^2) \bar{\beta}_{2n}}{(b\beta_{1n}^2 + H\beta_{1n}^2 \bar{\beta}_{2n} + b\bar{\beta}_{2n}^2 + Hb^2 \bar{\beta}_{2n}^3)} d\omega_n \quad (16)
 \end{aligned}$$

where the path of integration between  $i\omega + c_2 k_n$  and  $\omega_n^0$  satisfies the condition of  $q$  real and positive and  $k_n$  in the first quadrant, and between  $\omega_n^0$  and  $\infty$  proceeds along the real axis. As  $t \rightarrow r/c_2$ , the point of stationary phase approaches the point  $\omega_n^0$ , and an evaluation of the function there leads directly to (15).

In conclusion, we remark that the method developed here for the simple two-layer liquid model (or the analogous Love wave problem), has also been applied to problems involving more complicated systems of elastic layers, as well as to the case of asymptotic solutions valid for large values of the time variable. This method may be summarized as follows: the formal solution to the problem treated here is a function of two independent variables,  $\omega$  and  $k$ . We consider it as a function of these two independent variables from the very beginning, and therefore shift the branch line integral into the most rapidly converging position in both the  $k$  and the  $\omega$  planes. (This yields  $P_{B_1}$ .) In doing so, we pick up additional contributions from poles which were originally on the lower Riemann surface of the  $k$  plane. It is especially important to note that the dispersion curves of these poles provide a continuation of the normal mode dispersion curves past their cutoff values,  $\omega_n^0$ , to

complex values of  $\omega$  and  $k$ . (This is the contribution  $P_{B_2}$ .) For our final solution, we employ the method of stationary phase on these dispersion curves ( $P_{B_2} + P_N$ ) and also evaluate the rapidly converging repeated line integral ( $P_{B_1}$ ) asymptotically. We see that the pressure decay with horizontal range is of the order  $r^{-2}$ .  $P_{B_1}$  is a monotonically decreasing function of time, while  $P_{B_2} + P_N$  represent vibrations at the cutoff frequencies  $\omega_n^0$ .

*Acknowledgments*—The author wishes to thank Drs. A. Ginzburg, G. W. Postma, and E. Strick of the Shell Development Company for many helpful discussions.

#### REFERENCES

- KNUDSEN, W. C., Propagation of a Pressure Transient in a Two-Layer, Liquid Model, *J. Acoust. Soc. Am.*, **29**, 918-924, 1957.  
OFFICER, C. B., JR., The Refraction Arrival in Water Covered Areas, *Geophysics*, **18**, 805-819, 1953.  
PEKERIS, C. L., Theory of Propagation of Explosive Sound in Shallow Water, *Geol. Soc. Am. Mem.* no. 27, October 15, 1948.  
VAN DER CORPUT, J. G., Zur Methode der stationären Phase, II, *Compositio Mathematica*, **3**, 328-372, 1936.

(Manuscript received July 25, 1958.)

#### APPENDIX—COMPLEX DISPERSION CURVES

If we do not wish to ignore *a priori* poles in the complex  $k$  plane, we may proceed as follows. The original branch line  $\Gamma$  is drawn from  $\omega/c_2$  to the origin along the real  $k$  axis and then to  $-i\infty$  along the imaginary axis. The origin itself may be avoided by means of a small indentation in the fourth quadrant. There are now no poles of  $F_1(\omega, k)$  in the fourth quadrant of the top sheet of the  $k$  plane. We combine the two parts of the branch line integral, which refer to the two sides of the cut, in a manner analogous to that described in Section II above, and shift the path of integration to the position  $(BC)$  in Figure 1, that is, a straight line from  $\omega/c_2$  to  $\omega/c_2 - i\infty$ . In doing so, we may pass over poles of the integrand, which correspond to roots of the equation  $f_2 = 0$ . We pass over all roots,  $k_\nu$ , whose real parts are less than  $\omega/c_2$ ; those whose real parts are equal to  $\omega/c_2$  are avoided by means of a small indentation in the path of integration  $(BC)$  of Figure 1.

We introduce the new independent variable  $q$ ,

change order of integration, and write

$$P_B = -\frac{2A}{\pi} I m \int_0^\infty e^{-r q} dq \int_0^\infty \frac{e^{i\omega(t-r/c_2) + i\pi/4}}{(1/\theta + i\omega)} \cdot \left(\frac{2k}{\pi r}\right)^{\frac{1}{2}} F_2(\omega, q) d\omega \\ + 2A I m \sum_\nu \int_0^\infty \frac{e^{i(\omega_\nu t - k_\nu r) + i\pi/4}}{(1/\theta + i\omega_\nu)} \left(\frac{2k_\nu}{\pi r}\right)^{\frac{1}{2}} \frac{\sin \beta_{1\nu} z \sin \beta_{1\nu} d}{\sin \beta_{1\nu} H [\partial/\partial k (\beta_{1\nu} \cos \beta_{1\nu} H - b\bar{\beta}_{2\nu} \sin \beta_{1\nu} H)]} d\omega_\nu. \quad (A-1)$$

The first term on the right-hand side of expression (A-1) is identical with (2). In the second term, the partial derivative  $\partial f_2/\partial k$  in the denominator refers to constant  $\omega$ . The summation extends over  $\nu = 1, 2, \dots$  roots of the equation  $f_2 = 0$  in the complex  $k$  plane; that is, the equation  $f_2 = 0$  provides a multivalued relation (dispersion curves) between complex values of  $k$  and positive real values of  $\omega$ . The subscript  $\nu$  indicates evaluation of a function at such roots. The integrations over  $\omega_\nu$ , which are indicated to extend between the extreme limits 0 and  $\infty$ , actually depend on the complex dispersion curves. The condition that such a curve crosses the line  $(BC)$ , that is, that  $\text{Re}(k) = \omega/c_2$ , represents a limit of integration.

Evaluation of the first term of expression (A-1) should of course also provide an extension of the integrals beyond the limits just mentioned. This is analogous to the case of the real dispersion curves,  $n = 1, 2, \dots$ , discussed in the main part of the paper. The final integrals may be evaluated by van der Corput's method of critical points, and should yield pressure contributions which decay exponentially with  $r$ . Thus it is not worth while here to investigate the complex dispersion curves in detail. We may, however, sketch briefly the behavior of these curves for  $c_2/c_1 > 1$ .  $b < 1$ , in the fourth quadrant of the  $k$  plane. At  $\omega = 0$ ,  $k = (2H)^{-1} \ln [(1+b)/(1-b)] - i(\nu - \frac{1}{2})\pi/H$ , while  $dk/d\omega = 0$  and  $d^2k/d\omega^2 = (kc_1^2)^{-1} [1 + b(1 - c_1^2/c_2^2)(kH)^{-1}(1 - b^2)^{-1}]$ . These initial points are evidently on the lower Riemann surface\* and not in the shaded region

\*This refers to the equation  $f = \beta_1 \cos \beta_{1H} + b\bar{\beta}_2 \sin \beta_{1H} = 0$ . As was pointed out in the main part of the paper, for  $f_2 = \beta_1 \cos \beta_{1H} - b\bar{\beta}_2 \sin \beta_{1H} = 0$ , the definition of upper and lower Riemann surface is reversed.

of Figure 1. They arise thus in the evaluation of the integral  $P_{B_1}$  of (3), where the meaning of the summation over  $n$  may be extended to include the indices  $\nu$ . (It is readily seen that such initial points cannot occur in the first quadrant of the complex  $\omega$  plane for values of  $q \rightarrow 0$ ,  $\omega \rightarrow 0$ , or phase ( $\omega$ ) = 0 or  $\pi/2$ .) Returning to the  $k$  plane, we note that with increasing  $\omega$  the roots move toward the real axis and for a while may appear in the shaded region of Figure 1. They touch the real  $k$  axis at points (evidently not in the shaded region of Figure 1) which are described by the equation  $b^2 \tan^3 \beta_1 H - \beta_1 H \tan^2 \beta_1 H + \tan \beta_1 H - \beta_1 H = 0$ . At these points, the roots in the fourth quadrant of the  $k$  plane coincide with a similar set of roots in the first quadrant, and the period equation has double roots. With a further increase of  $\omega$ , there will now be two sets of roots on the real  $k$  axis. One set moves toward the branch point  $k = \omega/c_2$  and appears there on the upper Riemann surface as the normal mode poles,  $n = 1, 2, \dots$ ; the other set moves on the lower Riemann surface toward the limiting point  $k = \omega/c_1$  or  $(\omega/c_1) [(1 - b^2 c_1^2/c_2^2)/(1 - b^2)]^\dagger$ . This indicates the connection between the modes  $n$  and  $\nu$ .

#### ADDENDUM

In the review of this paper, the question has been raised as to whether it is permissible to express the solution to the problem directly in terms of contributions due to poles in the complex  $k$  and  $\omega$  planes, since such contributions may increase exponentially with the depth  $z$ . In contrast with the steady-state harmonic problem (see Ewing, Jardetsky and Press, *Elastic Waves*

in *Layered Media*, McGraw-Hill Book Co., New York, 1957, p. 136), where such an increase of signal amplitude with depth would lead to difficulties, no similar restriction on the form of the solution is required for the pulse problem. When a source starts emitting a signal at time  $t = 0$ , then we can show that there is no contribution to the final solution at a detector position  $(r, z)$  until the arrival of a first disturbance travelling by the shortest time path. The total signal is therefore limited to a finite region of the space-time domain. The complex poles represent progressive waves which radiate into the lower half space, and we may expect these waves to decrease exponentially with time. But such waves will also increase exponentially with distance (at constant time) in the direction of propagation and thus increase with depth. For the pulse problem, where the final solution is limited to a finite region of the space-time domain, it is quite correct to construct solutions by a superposition of waves decaying exponentially in time, even though such solutions individually are not proper for the steady-state problem for in the final integration the contributions ahead of the first arrival cancel.

A complete solution of the present problem shows that the complex pole contributions do indeed yield the proper response of the medium to a point source pulse. We remark that the formulation of the asymptotic solution in the present note was chosen to agree with the papers of *Pekeris* [1948] and *Officer* [1953] and is not very convenient for a discussion of the complex pole contributions.

(Addendum received October 6, 1958.)

# Calculations Based on the Kozeny-Carman Theory

WALTER D. ROSE

*Department of Mining and Metallurgical Engineering  
University of Illinois, Urbana, Illinois*

**Abstract**—Fluid flow in porous media, such as sands, sandstones, and carbonate rock, can be analyzed by experiment (Darcy's law) and by the Kozeny-Carman theory. The latter is based on the supposition that somehow the conductivity (permeability) of porous rocks is related to pore shapes and tortuosities, to the porosity, and to the boundary surface area against which the flowing viscous fluid is dragging. To test these ideas, a model of porous rock is selected which may be described as a random packing of spheres of various sizes. The porosity and surface-area terms are then calculated from geometric considerations. Then the character of the model is modified by assuming that a portion of the pore space is occupied by an immobile wetting phase located as pendular rings at each intersphere contact. The effective porosity and surface-area terms are calculated for this new condition, so that the effect of the pendular ring material on permeability can be computed, and so that a check of the Kozeny-Carman theory can be obtained.

The foregoing procedure is an indirect method which avoids the complication of explicitly evaluating pore shape and tortuosity parameters. It is shown that permeability appears to increase as pendular ring material is added to the packing, to an extent not inconsistent with expectations.

The physics of flow through porous media is of interest in many scientific and industrial investigations. Frequently experimental procedures do not give all the information that is needed, and a theoretical approach is indicated, to reveal the character of the mechanism involved. A problem of special importance, and special difficulty, has to do with the flow of a fluid which only partially occupies the interstices of the conducting porous medium. The Kozeny-Carman theory has been applied to this problem, and a geometric check of such an application is given below.

Many questions have been raised about the validity of Kozeny-Carman theory as providing a correct description of fluid flow through porous media. Scheidegger [1957], for example, cites the popular objections. The textural constant and specific surface-area terms to be considered are frequently indeterminate both as to definition and as to measurement; hence, direct experimental verifications usually are not possible. The usefulness of the Kozeny-Carman approach, nonetheless, can be examined to some extent by calculations based on a consideration of fluid flow through simplified models of porous media.

Let us examine the properties of a random packing of spheres of random size. If we are to accept the Kozeny-Carman theory, then

$$K = f/tA^2 \quad (1)$$

where  $K$  is the permeability,  $f$  is the fractional porosity, and  $A$  is the specific surface area of the pore walls per unit pore volume. When part of the pore space is occupied by a wetting phase located as pendular rings at every point of intersphere contact, in general

$$K_e = f_e/t_e A_e^2$$

or

$$K_{rn} = S_n(A/A_e)^2(t/t_e) = K_e/K \quad (2)$$

where  $K_e$  is the permeability afforded to the flow of the nonwetting fluid when pendular rings of wetting fluid occupy a portion of the pore space,  $S_n$  is the fraction of the pore space not occupied by these pendular rings of wetting fluid, and  $A_e$  is the specific surface area per unit pore volume forming the boundaries of the nonwetting fluid flow channels. The latter will include that portion of  $A$  not covered by the pendular rings of wetting fluid, plus whatever fluid-fluid interfacial surface area exists between the wetting and nonwetting phases.

The  $t$  and  $t_e$  terms appearing in (1) and (2) are the relevant textural constants of proportionality which, according to the Kozeny-Carman theory, are related to pore shapes and tortuosities. How these terms are to be numerically evaluated is a question of major importance, heretofore not fully answered, in spite of the



reports [Scheidegger, 1957] that (1) tortuosity is independently determined by an electrical resistivity measurement, and (2) the pore shape factor has a universal value close to 2.5 in a way which (paradoxically enough) is independent of pore shape.

In the present analysis  $t/t_e$  is assumed to be close to unity (perhaps within 99 pct), since we choose to consider only those systems where small amounts of pendular-ring wetting fluid are added to the sphere packing. This is a device to make unnecessary the almost impossible task of explicitly evaluating  $t$  and  $t_e$  for complex sphere packings. (Dalla Valle [1943] has done this for a simple case of closely packed spheres of uniform size.) Thus, we may imagine that pore shapes and tortuosities affecting the flow of the nonwetting fluid are only negligibly altered by the presence of small amounts of pendular-ring wetting fluid, in the sense that the fluid particles of the nonwetting fluid generally flow through openings and in directions little influenced by whether or not small pendular rings of wetting fluid exist within the porous continuum.

Thus, by limiting our attention to the special case where  $t/t_e$  can be neglected, we find the problem of applying the Kozeny-Carman theory greatly simplified. In fact, for the random packing of random-sized sphere model, we need only to know something about the volumes and surface areas of the pendular rings of wetting fluid in order to solve (2) by calculation. Thus, by simple geometric considerations, we are enabled to obtain a check of the Kozeny-Carman theory.

The volumes and surface areas of pendular rings have been derived elsewhere [Rose, 1958] and have the form (see Figs. 1 and 2)

$$v_{ij} = F_1(\theta_i, r_i, r_j) \quad (3)$$

and

$$a_{ij} = F_2(\theta_i, r_i, r_j) \quad (4)$$

where  $r_i$  is the radius of the larger, and  $r_j$  is the radius of the smaller of any two spheres in contact;  $\theta_i$  is the angle between the line joining the sphere centers and that joining the center of the larger sphere with the edge of the pendular ring; and  $F_1$  and  $F_2$  are complicated trigonometric functions based on approximation of the surface of the pendular rings by arcs of circles instead of by catenaries.

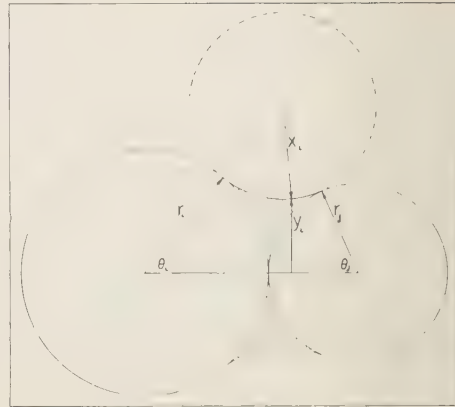


Fig. 1—A depiction of the  $ij$  contact between two spheres of radius  $r_i$  and  $r_j$  showing the pendular ring defined by the angles  $\theta_i$  and  $\theta_j$ , and by the principle radii of curvature  $x_i$  and  $y_i$

Figure 2 shows that (3) and (4) are linear functions, having the form:

$$v_{ij} = r_i^3(\theta_i/G)^H \quad (3a)$$

$$a_{ij} = r_i^2(\theta_i/M)^N \quad (4a)$$

where the following values of the constants hold:

$r_i/r_j$	$G$	$H$	$M$	$N$
1.0	80	3.43	30	2.86
0.5	66	3.48	28	2.81
0.1	53	3.55	25	2.75

Intermediate values, as may be needed to solve (7) and (9) below, are obtainable by interpolation.

Therefore, to solve (2), note that

$$A = (4\pi/f) \sum n_i r_i^2 \quad (5)$$

and

$$A_e = A + A_P - A_D \quad (6)$$

where

$$A_P = (1/f) \sum n_{ij} a_{ij} \quad (7)$$

$$A_D = (2\pi/f) \sum n_{ij} [(1 - \cos \theta_i) r_i^2 + (1 - \cos \theta_j) r_j^2] \quad (8)$$

and where  $f$  is the fractional porosity;  $n_i$  is the number of  $r_i$  spheres per unit bulk volume;  $n_{ij}$  is the number of contacts between  $r_i$  and  $r_j$  spheres per unit bulk volume; and  $\theta_i$  is the

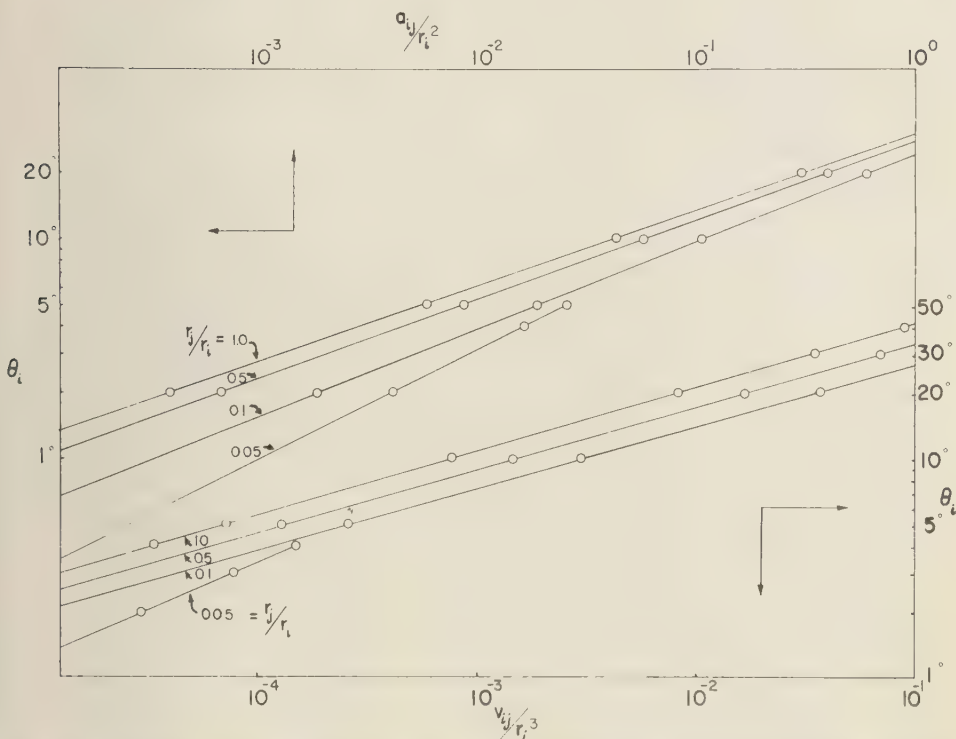


Fig. 2—A graphical solution of equations 3 and 4, correcting minor computational errors previously given by Rose [1958]

angle in the  $r_j$  sphere corresponding to  $\theta_i$  in the  $r_i$  sphere, and is obtained from

$$\tan(\theta_j/2) = [r_i/r_j] \tan(\theta_i/2)$$

Finally,  $S_n$  in (2) is given by

$$S_n = 1 - (1/f) \sum n_{ij} v_{ij} \quad (9)$$

Therefore, upon neglecting  $t/t_s$  and applying (3) through (9), values of nonwetting phase relative permeability can be calculated by (2) as soon as some assumption is made about the size of the pendular rings to consider. Three possible approaches are listed below in the order of increasing relevancy.

(A) Assume a constant value of  $\theta_i$  for each  $r_i$  sphere contact [Rose, 1957ab, 1958]. This is reasonable (and convenient) when dealing with packings of spheres of uniform size, or when considering consolidation of sands where the deposition is independent of the sphere size ratio.

(B) Let the curvature  $C_i$  of each pendular ring be constant, that is, uniformly the same throughout the porous medium, but avoid any pendular ring so large that it will coalesce with contiguous pendular rings. This represents, for example, the spatial distribution of connate water in a water-wet (virgin) petroleum reservoir, as it is found at points far removed from the water-oil contact. The implicit assumption is being made here that geologic time is sufficiently long for static equilibrium to be reached via diffusion, so that transfer need not depend upon continuity between pendular rings.

(C) Again let the curvature of each pendular-ring interface have some constant value, but impose the restriction that the wetting phase must be everywhere continuous. In this case, some of the pendular rings would necessarily coalesce with contiguous pendular rings, and the spatial distribution of wetting and nonwetting fluids would be in the funicular regime.

TABLE 1—Calculated values of  $K_{rn}$  for 'filled' rhombohedral packings, with  $C_i = 100$ 

Parameter	Value						
1a.—Packing components							
Code designation of sphere size (after <i>White and Walton</i> , 1937)	<i>E</i>	<i>J</i>	<i>K</i>	<i>L</i>	<i>M</i>		
relative sphere size ( $r_i, r_j$ 's)	1.000	0.414	0.225	0.177	0.116		
$n_i$ (unit bulk volume = 5.66)	0.177	0.177	0.354	1.415	1.415		
1b.—Character of contacts							
Code identification of contacts	<i>EE</i>	<i>EJ</i>	<i>EK</i>	<i>EL</i>	<i>JL</i>	<i>EM</i>	<i>KM</i>
$r_j/r_i$	1.000	0.414	0.225	0.177	0.427	0.116	0.516
$n_{ij}$	1.062	1.062	1.415	4.245	1.415	1.415	1.415
$\theta_i$ (degrees)	7.8	5.9	4.7	4.3	5.9	3.8	6.4
$\theta_j$ (degrees)	7.8	14.1	20.4	24.0	13.5	32.0	12.2
$a_{ij}/r_i^2$	0.021	0.016	0.0115	0.0112	0.015	0.0092	0.016
$v_{ij}/r_i^3$	$3.3 \times 10^{-4}$	$2.5 \times 10^{-4}$	$1.5 \times 10^{-4}$	$1.3 \times 10^{-4}$	$2.5 \times 10^{-4}$	$9.0 \times 10^{-5}$	$2.8 \times 10^{-4}$
1c.—Character of packings							
Sphere sizes in packing	<i>E</i>	<i>E, J</i>	<i>E, J, K</i>	<i>E, J, K, L</i>		<i>E, J, K, L, M</i>	
Sphere contacts in packing	<i>EE</i>	<i>EE, EJ</i>	<i>EE, EJ, EK</i>	<i>EE, EJ, EK, EL, JL</i>		<i>EE, EJ, EK, EL, EM, JL, KM</i>	
Resultant packing porosity, <i>f</i>							
	0.2595	0.207	0.190	0.158		0.149	
<i>A</i>	8.557	12.573	14.881	21.415		24.312	
<i>A<sub>p</sub></i>	0.086	0.190	0.293	0.677		0.846	
<i>A<sub>D</sub></i>	0.475	0.935	1.328	2.621		3.073	
<i>A<sub>e</sub></i>	8.168	11.828	13.846	19.471		22.085	
( <i>A/A<sub>e</sub></i> ) <sup>3</sup>	1.098	1.130	1.156	1.210		1.212	
<i>s<sub>n</sub></i>	0.9987	0.9970	0.9957	0.9911		0.9896	
<i>K<sub>rn</sub></i> ( <i>C<sub>i</sub></i> = 100)	1.097	1.127	1.151	1.199		1.198	
<i>K<sub>rn</sub></i> [ <i>Rose</i> , 1957b for $\theta_i = 10^\circ$ ]	1.145	1.227	1.315	1.639		1.710	

Admittedly, transfer via diffusion is not always an important aspect of many multiphase fluid-flow processes; therefore, we would like to investigate (2) for funicular saturation conditions. No simple way is known to describe the geometry of funicular interfacial boundaries when dealing with random packings of random-sized spheres, and one would have the added complication of taking into account transfer of viscous forces at fluid-fluid interfaces [Yuster, 1951]. Hence, at this time, it appears possible only to make calculations for systems in the pendular regime, for which we can employ (3) through (9) directly if we are willing to neglect the effect of finite velocities at interfacial boundaries. The latter approximation is justified to some extent by the observation that transfer of viscous forces at fluid-fluid interfaces can be neglected when the interfacial surface area is small and when there is no continuity between contiguous pendular rings of wetting fluid [Scott and Rose, 1953].

The curvature  $C_i$  referred to above are simply the sum of the reciprocals of the principal radii of curvature, or

$$C_i = (x_i + y_i)/x_i y_i \quad (10)$$

where

$$x_i = \frac{r_i[(r_i + r_j) - (1 + r_j) \cos \theta_i]}{(r_i + r_j) \cos \theta_i - (r_i - r_j)}$$

$$y_i \cong [(r_i + x_i) \sin \theta_i] - \left[ x_i \cos \frac{(\theta_i - \theta_j)}{2} \right]$$

Thus, let us first consider a 'filled' rhombohedral packing of spheres [White and Walton, 1937] in the manner previously treated by Rose [1957b, 1958] but instead of assuming a constant value for  $\theta_i$  for each pendular ring, we now postulate a constant value of  $C_i = 100$  (see Approach 2, above). Table 1 and Figure 3 show the results which are obtained, and it is again indicated that the nonwetting phase relative permeability (as calculated by (2)) increases with decreasing values of  $S_n$ . The

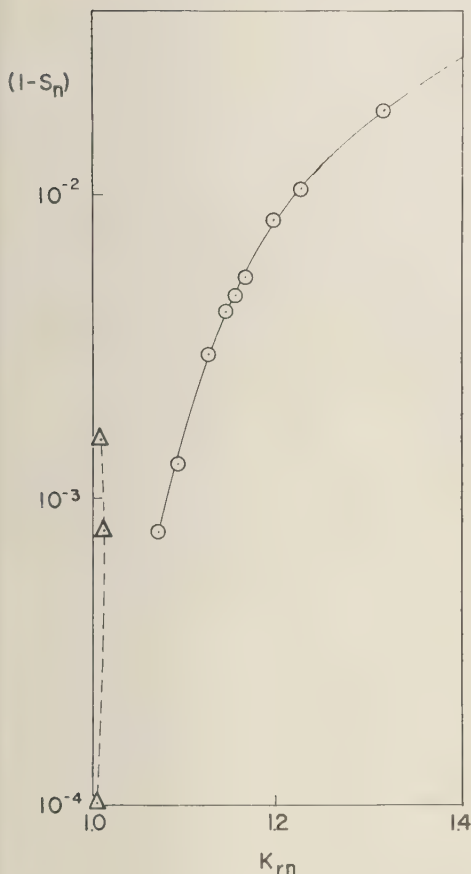


FIG. 3—Nonwetting-phase relative permeability  $K_{rn}$  as a function of the fraction of the pore volume occupied by the pendular-ring material  $(1 - S_n)$ ; the solid curve is from the data of Rose [1957b] for the case of a constant value of  $\theta_i = 10^\circ$ ; the circles are the data of Table 1, and the triangles are the data of Table 2, both for the case of constant value of  $C_i = 100$

magnitude of this anomalous increase is not nearly as great as heretofore reported [Rose, 1957b], in part because of the way the  $S_n$  values have been minimized, but calculated  $K_{rn}$  values are still great enough to perpetuate questions about the validity of Kozeny-Carman theory. In effect, it appears that the Kozeny-Carman theory gives too much weight to the surface area of the pore walls located at the distant points where there is negligible flow.

Already demonstrated is the proposition that

geometry requires  $A_s$  to decrease when pendular ring material is added to a sphere packing (see (6)) in such a way that  $K_{rn}$ , as calculated by (2), always increases as  $S_n$  first decreases. The original assumption of a constant  $\theta_i$  accentuated this effect, and the present assumption of a constant  $C_i$  minimizes (but does not destroy) the effect. Hence, we still seem to have the paradox of theory predicting an increase in  $K_{rn}$  when intuition suggests a decrease. Unless and until the question is resolved, Kozeny-Carman theory perhaps deserves the status of reserved acceptance.

Careful experimental work can remove the uncertainty. In the absence of the requisite data one can only speculate on the significance of the above findings. It can be noted, for example, that almost all relative permeability data for nonwetting phases show [Scheidegger, 1957] that  $K_{rn}$  reaches unity when  $S_n$  is still less than unity; moreover, the extant data seldom pertain to saturation conditions in the interval between the point at which  $K_{rn}$  first becomes unity and the point at which  $S_n = 1$ . On the other hand, early results of unknown significance given by Hassler [1938] show values of  $K_{rn}$  greater than unity in this interval, and Yuster [1951] implies the existence of similar data. Hence, we may say that experimental results do not yet disprove the indications of (2) (that is, that  $K_{rn}$  will increase as portions of the pore space are first filled by pendular rings of wetting fluid), and, we may wonder if these increases have in fact been observed, but not reported, in the same way that suspect gas permeability results were observed but not reported until a theoretical explanation of gas-slippage had been developed [Klinkenberg, 1941]. The proof of this speculation must obviously await the completion of careful experimental work.

Improved correspondence between theory and expectations can be achieved in another way, however. Eq. (3) through (10) are perfectly general for random as well as regular packings of spheres. Hence, we need not limit our attention to the regular arrangements discussed by Dalla Valle [1943], White and Walton [1937] and Rose [1957ab, 1958]. Indeed, it is our desire to choose values of  $r_i$ ,  $r_j$ ,  $n_i$ , and  $n_{ij}$  characteristic of the sand packings which occur in nature, to achieve a reasonable model of the prototype systems.



TABLE 2—Calculated values of  $K_{rn}$  for an arbitrary packing of three sphere sizes, with  $C_i = 100$ 

Parameter	Value					
1a—Packing Components						
Code designation of sphere size	<i>E</i>	<i>J</i>	<i>K</i>			
relative sphere size ( <i>r<sub>i</sub></i> , <i>r<sub>j</sub></i> 's)	1.000	0.500	0.100			
<i>n<sub>i</sub></i> (unit bulk volume = 15)	0.133	0.333	1.000			
1b—Character of Contacts						
Code identification of contacts	<i>EE</i>	<i>EJ</i>	<i>JJ</i>	<i>EK</i>	<i>JK</i>	<i>KK</i>
<i>r<sub>j</sub>/r<sub>i</sub></i>	1.000	0.500	1.000	0.100	0.200	1.000
<i>n<sub>ij</sub></i>	0.067	0.667	0.333	2.000	3.333	1.000
<i>θ<sub>i</sub></i> (degrees)	7.8	6.2	7.8	3.6	4.5	7.8
<i>θ<sub>j</sub></i> (degrees)	7.8	12.5	7.8	34.9	22.2	7.8
<i>a<sub>ij</sub>/r<sub>i</sub><sup>2</sup></i>	0.021	0.0145	0.021	0.0083	0.012	0.021
<i>v<sub>ij</sub>/r<sub>i</sub><sup>3</sup></i>	$3.3 \times 10^{-4}$	$2.7 \times 10^{-4}$	$3.3 \times 10^{-4}$	$7.7 \times 10^{-5}$	$1.5 \times 10^{-4}$	$3.3 \times 10^{-4}$
1c—Character of packings						
Sphere sizes in packing	<i>E</i>	<i>E,J</i>	<i>E,J,K</i>			
Sphere contacts in packing	<i>EE</i>	<i>EE,EJ,JJ</i>	<i>EE,EJ,EK,JJ,JK,KK</i>			
Resultant packing porosity, <i>f</i>	0.442	0.268	0.263			
<i>A</i>	3.791	21.882	70.159			
<i>A<sub>p</sub></i>	0.003	0.048	0.150			
<i>A<sub>D</sub></i>	0.018	0.251	0.558			
<i>A<sub>s</sub></i>	3.776	21.679	69.751			
$(A/A_s)^2$	1.008	1.019	1.012			
$(1 - s_n)$	$9.9 \times 10^{-5}$	$8.1 \times 10^{-4}$	$1.7 \times 10^{-3}$			
<i>K<sub>rn</sub></i>	1.007	1.018	1.010			

While the particle-size distribution is derivable from sieve analysis, the  $n_{ij}$  values are not so easily obtained. *Smith* and others [1929] describe a novel technique for preferentially etching the contacts in random packings of lead spheres of uniform size so that the contacts could be counted. This technique is certainly laborious, and perhaps not easily adaptable to a consideration of the sand packings as found in nature. Hence, other methods must be devised to evaluate the frequency of occurrence of particle contacts  $n_{ij}$ , such as certain procedures for investigating thin sections which are available for the treatment of consolidated sandstones.

On the other hand, it is still possible to investigate the consequence of assuming various particle-size distributions and frequency of particle contact occurrence, relating these assumptions to the values of  $K_{rn}$  as calculated by (2). Table 2 describes a hypothetical packing made up of spheres of three sizes, where bias has been introduced in the following ways: (A) The frequency of occurrence of small particles is emphasized. Hence, the value of *A* in (1), (2), and (5) largely depends on the prevalence of small particles. (B) The frequency of occurrence of large to small sphere contacts is emphasized.

Hence  $A_p$  (the added interfacial specific surface area per unit pore volume) and  $A_D$  (the subtracted sphere 'dome' specific surface area per unit pore volume) are correspondingly influenced by the large  $n_{ij}$  values of the low  $r_j/r_i$  ratio contacts.

Geometry (see (3) and (4)) shows that as the  $r_i/r_j$  ratio increases,  $A_p$  approaches (and may exceed)  $A_D$ ; hence,  $A_s$  approaches *A* in a way which minimizes the values of  $K_{rn}$  calculated by (2). Also note the fact that the  $\theta_i$  values (and hence,  $a_{ij}$  and  $v_{ij}$  values) are minimized as the  $r_j/r_i$  ratio is decreased and the  $C_i$  values are held constant. This latter effect shown in Figure 4 also tends to minimize the  $K_{rn}$  values calculated by (2).

As might be expected, therefore, Table 2 and Figure 3 show calculated values of  $K_{rn}$  only slightly greater (one per cent) than unity as applicable to the selected model. Indeed, for this case one can assert that excellent correspondence between theory (2), expectations, and extant laboratory data has been achieved.

The above result has been obtained by saying that natural sand packings are characterized by randomness and a high frequency of occurrence of large to small particle contacts. This assump-

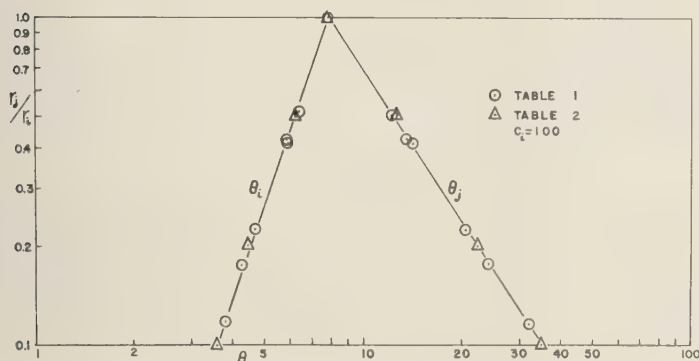


FIG. 4—The relationship between the angles  $\theta_i$  and  $\theta_j$  with  $r_i/r_j$  for a constant value of  $C_i = 100$

tion is by no means unreasonable when we realize that the existence of small particles in a given packing can only serve to diminish the occurrence of contacts between large particles and other nearby large particles.

In contrast, the paradox of calculating increasing  $K_{rn}$  with decreasing  $S_n$  had been reported earlier for regular packings of uniform-size spheres [Rose, 1957a], and for regular ('filled') packings of various size spheres [Rose, 1957b, 1958]. These systems, appear to be unsatisfactory models of natural sand packings, although superior to the bundle-of-capillary-tube models others have discussed [Scheidegger, 1957; Yuster, 1951].

As the matter stands, therefore, one might conclude that the relevancy of the Kozeny-Carman approach is by no means vitiated by the various analytic checks of (2) given previously [Rose, 1957ab, 1958]. Rather, it would be fairer to say that further calculations must be undertaken as soon as improved assumptions can be made about the  $n_{ij}$  values, and that carefully obtained experimental results are needed before firm conclusions can be drawn. In fact, Kozeny-Carman theory seems to be validated in part by the indications of Table 2 and their correspondence to available experimental data (although some of the fundamental objections of Scheidegger [1957]) still need to be investigated.

In any case, it is clear that (3) through (10) provide a novel means for specifying the characteristics of sand packings, and this would seem

to have a utility not solely dependent on the validity of Kozeny-Carman theory. It also appears that capillary pressure equilibrium (that is,  $C_i = \text{constant}$ ) is a reasonable and useful assumption to make in treating the various problems of mixture flow and interspatial fluid distribution in porous media. For random sphere models, Figure 4 shows an interesting and useful relationship between  $r_i/r_j$  and  $\theta_i$  (and  $\theta_j$ ) for conditions of constant  $C_i$ .

*Acknowledgment*—The perseverance of Don Hise in making the calculations summarized in Tables 1 and 2 is appreciated.

#### REFERENCES

- DALLA VALLA, J. M., *Micromeritics*, 428 pp., Pitman Pub. Co., New York, 1943.
- HASSLER, G. L., The measurement of permeability of reservoir rock and its application, *The Science of Petroleum*, 1, 198-208, Cambridge U. Press, 1938.
- KLINKENBERG, L. J., The permeability of porous media to liquids and gases, *Drilling and Production Practice*, pp. 200-213, American Petroleum Institute, 1941.
- ROSE, WALTER, Fluid flow in petroleum reservoirs, I, The Kozeny paradox, *Ill. State Geol. Surv. Circ. 237*, 8 pp., 1957a.
- ROSE, WALTER, Fluid flow in petroleum reservoirs, II, Effects of sand consolidation, *Ill. State Geol. Surv. Circ. 242*, 14 pp., 1957b.
- ROSE, WALTER, Volumes and surface area of pendular rings, *J. Appl. Phys.*, 29, pp. 687-691, 1958.
- SCHIEDEGGER, A. E., The physics of flow through porous media, MacMillan, 236 pp., 1957.
- SCOTT, P. H., AND WALTER ROSE, An explanation of the Yuster effect, *Trans. Amer. Inst. Min. Metal Eng.*, 198, 323-324, 1953.

- SMITH, W. O., P. D. FOOTE, AND P. F. BUSANG, Packing of homogeneous spheres, *Phys. Rev.*, **34**, 1271-1271, 1929.
- WHITE, H. E., AND S. F. WALTON, Particle packing and particle shape, *J. Amer. Ceramic Soc.*, **20**, 155-163, 1937.
- YUSTER, S. T., Theoretical considerations of multiphase flow in idealized capillary systems, *Proc. Third World Petroleum Congress*, sec. 2, pp. 436-445, 1951.
- (Communicated manuscript received January 23, 1958, and, as revised, August 12, 1958.)

# Systematic Determination of Unit Hydrograph Parameters

J. E. NASH

*Hydraulics Research Station  
Howbery Park, Wallingford, Berks., England*

**Abstract**—In unit hydrograph correlations the number of degrees of freedom which it is useful to maintain in the form of the instantaneous unit hydrograph is shown to be limited by the number of significant independent correlations with the catchment characteristics. The moments of the instantaneous unit hydrograph are suggested as a series of parameters of the response, for which correlations should be sought. A simple method of obtaining these moments is evolved, and a method of choosing, in any given case, between several two-parameter forms is demonstrated.

**Notation**—The following notation is used:

- IUH—instantaneous unit hydrograph
- $i$ —input (effective rainfall) a function of time
- $u$ —the indicial response (instantaneous unit hydrograph) a function of time
- $T$ —duration of inflow
- $V$ —storage
- $K$ —a linear storage parameter
- $s$ —output (storm runoff), a function of time
- $a, b, c$ —times from the origin to the centers of area of the input, the output, and the indicial response respectively
- $\tau, \epsilon, \gamma$ —time measured from the centers of area of input, indicial response, and output, respectively
- $I_n, U_n, S_n$ —the  $n$ th moments about the centers of area of the input, indicial response, and output, respectively
- $I'_n, U'_n, S'_n$ —corresponding moments about the origin
- $m_n = U_n/c^n$ —the  $n$ th dimensionless moment of the indicial response
- $h, g$ —parameters in the equation of the log-normal frequency distribution

**Number of parameters**—The essence of the unit hydrograph theory is the assumption of linearity in the relation between storm runoff and effective rainfall. Having made this assumption, it remains to establish the relation between the characteristics of the catchment and the response in the storm runoff to a predetermined 'input' of effective rainfall, that is, the indicial response. The most frequently used indicial response is the IUH, defined as the hydrograph of storm runoff caused by unit volume of effective rainfall, generated instantaneously and uniformly, over the catchment. To establish the relationship

between the IUH and the catchment characteristics by means of a statistical correlation, it is necessary to define the IUH and the catchment characteristics by numerical parameters. The parameters of the IUH may be actual measurements of, or constants in, the assumed equation of the IUH. Correlations must be established between the IUH parameters as dependent variables and the catchment characteristics as independent variables.

Such correlations may be classified as 'one parameter,' 'two parameter,' . . . depending on the number of parameters of the IUH for which independent correlations are obtained. Obviously, the greater the number of parameters we succeed in correlating the better, as we are thus enabled to use a more general IUH form.

If we measure only a single parameter, then we must assume that all IUH's having equal values of this parameter are identical. If we use two parameters, we can obtain a better fit to an actual IUH, due to the greater generality of the two-parameter form. It is not necessary, though it is desirable, that the form of IUH be expressed as an actual algebraic equation  $u = u(t)$  (discharge as a function of time). For example, one could measure the peak and assume that the shape is given by Commons [1942] basic hydrograph (one parameter), or one could assume that the IUH was generated by routing an isosceles triangular inflow of base  $T$  hours through linear reservoir storage,  $V = Ks$  (two parameters).

Clearly, the greater generality of a many-parameter form can be availed of, only if significant independent correlations are obtained for each parameter. In the two-parameter case mentioned (namely, where the IUH is assumed to be generated by routing an isosceles triangular



inflow of base  $T$  hours through linear reservoir storage), if correlations for  $K$  and  $T$  are obtained with different combinations of catchment characteristics we have a genuine two-parameter correlation. However if both  $T$  and  $K$  are correlated with the same catchment characteristics in such a manner that the catchment characteristics can be eliminated between the two equations and  $K$  expressed as a function of  $T$ , then we have only a one-parameter correlation because  $K$  can be replaced by the corresponding function of  $T$ .

Intermediate between these two cases we might find  $K$  and  $T$  correlated with almost the same catchment characteristics. In such a case it might be difficult to say whether we had achieved a one- or two-parameter correlation.

There are two precautions which help to prevent this situation arising. We may take as our two parameters  $T$  and  $T/K$  which are much less likely to be mutually correlated than  $T$  and  $K$ . We may also include the first parameter among the independent variables when studying the regression of the second parameter on the catchment characteristics. If a multiple linear regression is used, the values of the  $\beta$  coefficients [Ezekiel, 1941] show the relative proportions of the variance of the second dependent variable which is attributable to variations in the first parameter and to variations in the catchment characteristics themselves.

*Moments as parameters*—In defining the IUH parameters to use in the correlation we must remember that the parameters must be evaluated, by some curve-fitting process, from the records for each catchment used in the correlation. Therefore the parameters must be stable, in that small errors in the records must not produce large variations in the parameters.  $T$  and  $K$  of the routed isosceles triangle are not stable in this sense. Obviously a fairly large change in  $T$  and  $K$  in opposite directions would produce only a small change in the IUH generated, and conversely. In fact if we wished to use this particular two-parameter form we should correlate  $(T + K)$  and  $T/K$  with the catchment characteristics.  $(T + K)$  could be obtained from the records with greater certainty than either  $T$  or  $K$ . The second point to be remembered is the ease or difficulty with which the parameters may be evaluated from the records. For example, if we use the peak of the IUH as a parameter, it is necessary to derive from the records the actual

IUH for each catchment, a very difficult task.

It is the author's suggestion that the form of the IUH should not be determined until after the correlation, and that the moments of the IUH, which can be obtained very easily from the records without deriving the actual IUH, should be used as the parameters of the response. It is shown in Appendix 1 that the first moment of the IUH about the instant of effective rainfall is equal to the difference between the first moments of the storm runoff and effective rainfall about the time of beginning of effective rainfall. It is also shown that the second and third moments of the IUH about its center of area, are respectively equal to the differences between the corresponding moments of storm runoff and effective rainfall, each about its own center of area. The corresponding relation for higher moments is also derived.

Thus, to obtain as many parameters of the IUH as we wish, we simply calculate the moments of the storm runoff and effective rainfall and apply (1) and (4) of Appendix 1.

To obtain independence between the parameters of the IUH it is better to use, instead of the  $n$ th moment, the ratio of the  $n$ th moment to the first moment in the power of  $n$ . This renders all the parameters except the first dimensionless. In the notation defined above, the parameters become  $c, m_2, m_3, \dots$ . These parameters should be derived for each catchment and treated as dependent variables. Having obtained a correlation between  $c$  and the catchment characteristics, a second correlation should be sought between  $m_2$  as dependent variable, and  $c$  and the catchment characteristics as independent variables, and so on.

At some value of  $n$  it will be found that  $m_{n+1}$  shows no significant correlation with the catchment characteristics. It is possible that it might show significant correlation with  $c, m_2, m_3, \dots, m_n$ , but this is unimportant at this stage. We would thus have achieved an  $n$  parameter correlation and we must then seek an  $n$  parameter form for the IUH. To decide between various forms we might use the chi-squared test, used by statisticians to measure the closeness of fit of a frequency distribution. If  $n$  is small, as unfortunately is most probable, a method similar to that shown in Appendix 2 might be used to judge the suitability of various two-parameter forms. Here the ability of each form to predict the

third moment, when its two parameters are determined by equating first and second moments, is used as a test of suitability.

Concerning the number of parameters for which we might hope to obtain correlations it may be noted that the correlations of *Bernard* [1935], *McCarthy* [1940], *Snyder* [1938], and *O'Kelly* [1955] are all 'one parameter' while that of *Taylor and Schwartz* [1953] is 'two parameter.'

#### APPENDIX 1—THE EFFECT OF A LINEAR TRANSFORMATION ON MOMENTS

The object here is to establish the relationship between the moments of the 'input,' the 'indicial response' to a unit pulse, and the 'output.' Let us assume that the input and output are expressed in units which make  $\int i \, d\tau$  and  $\int s \, d\gamma$  both unity. This does not limit the generality of the equations subsequently derived, but it simplifies the expressions slightly.

In the transformation, the elementary strip  $i \, d\tau$  (Fig. 1), is replaced by an elementary output whose center of area must be later than  $\tau$  by  $c$ . This applies to every elementary strip; therefore the center of area of the output must be later than that of the input by  $c$ .

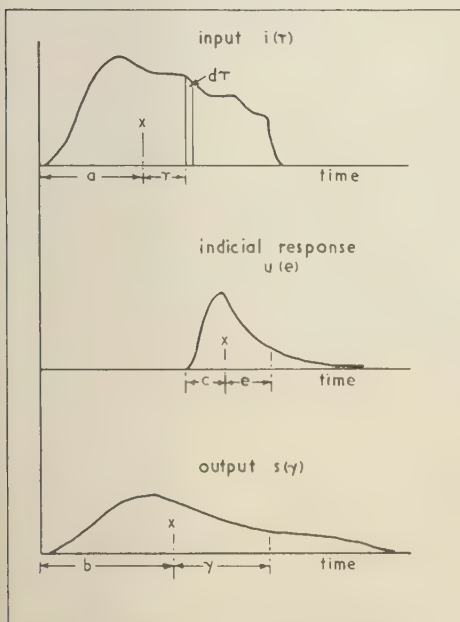


FIG. 1—The linear transformation

$$b = a + c \quad (1)$$

This is the relation between the first moments about the origin. The relation between the  $n$ th moments about the centers of area is found as follows.

Du Hamel's integral gives the output at time  $\gamma$  as

$$s = \int i(\tau)u(\epsilon) \, d\tau \quad \text{where} \quad \gamma = \epsilon + \tau.$$

$$\therefore S_n = \iint i(\tau)u(\epsilon)(\epsilon + \tau)^n \, d\tau \, d\epsilon \quad (2)$$

which may be expanded as

$$S_n = \iint i(\tau)u(\epsilon)(\epsilon^n + n\epsilon^{n-1}\tau + \frac{n(n-1)}{2!}\epsilon^{n-2}\tau^2 + \dots + \tau^n) \, d\epsilon \, d\tau.$$

Now

$$U_n = \int u(\epsilon)\epsilon^n \, d\epsilon, \quad \text{and} \quad I_n = \int i(\tau)\tau^n \, d\tau.$$

$$\begin{aligned} \therefore S_n &= \int i(\tau)(U_n + nU_{n-1}\tau + \frac{n(n-1)}{2!}U_{n-2}\tau^2 + \dots + \tau^n) \, d\tau. \\ \therefore S_n &= U_n + nU_{n-1}I_1 + \frac{n(n-1)}{2!}U_{n-2}I_2 + \dots + I_n. \end{aligned} \quad (3)$$

If we write the suffixes as power indices (without of course interpreting them as such, except for the purpose of expansion), we get

$$S_n = (U + I)^n. \quad (4)$$

This equation is quite general. It has very simple expansions for small values of  $n$ . Remembering that  $U_1 = I_1 = S_1 = 0$  we get

$$S_2 = I_2 + U_2 \quad (5a)$$

$$S_3 = I_3 + U_3 \quad (5b)$$

$$S_4 = I_4 + U_4 + 6I_2U_2 \quad (5c)$$

Eq. (4) and (1) enable us to calculate the moments of the IUH very simply from the moments of the storm runoff and effective rainfall.

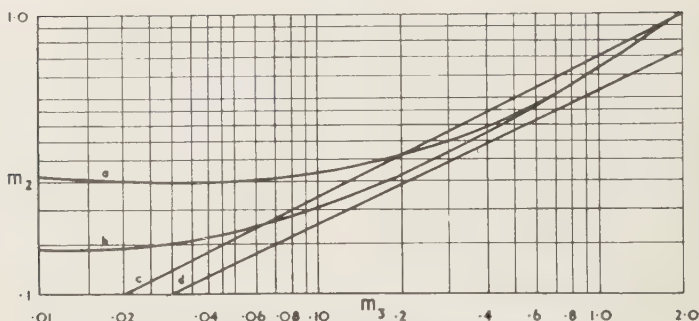


FIG. 2—Comparison of certain two-parameter forms

#### APPENDIX 2—CALCULATION AND COMPARISON OF THE MOMENTS OF CERTAIN TWO-PARAMETER FORMS

Here the quantities  $c$ ,  $m_2$ , and  $m_3$  are calculated for a number of possible two-parameter IUH forms. The relationship between  $m_3$  and  $m_2$  is calculated and plotted as a set of curves (Fig. 2). If we obtain, as described earlier in this paper, the values of  $m_2$  and  $m_3$  for each of several catchments, we can decide which of the two-parameter forms gives the best representation of the several IUH's by plotting corresponding values of  $m_3$  and  $m_2$  on Figure 2. The form corresponding to the line which passes closest to the scatter so obtained is the most suitable form.

The routed rectangle (Fig. 2, Curve a)—Figure 3

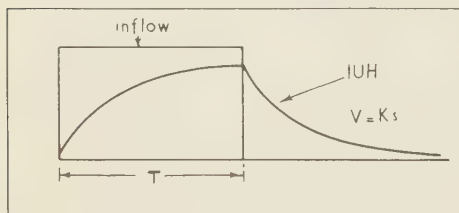


FIG. 3—The routed rectangle

shows the method of generation of this form. A rectangular inflow of duration  $T$  is routed through linear storage  $V = Ks$ .

The process of routing through linear storage is a linear transformation, and therefore the equations developed in Appendix 1 may be used to obtain the moments of the routed rectangle. In this case the response to a 'unit pulse' input is given by  $u = e^{-t/K}/K$

$$U_n' = \int_0^{\infty} \frac{t^n}{K} e^{-t/K} dt$$

$$= K^n \int_0^{\infty} \left(\frac{t}{K}\right)^n e^{-t/K} d(t/K) = K^n \Gamma(n)$$

whence  $c = K$ ,  $U_2 = K^2$ , and  $U_3 = 2K^3$ . The moments of the input are  $a = T/2$ ,  $I_2 = T^2/12$ , and  $I_3 = 0$  whence by (1) and (5),  $b = T/2 + K$ ,  $S_2 = K^2 + T^2/12$ , and  $S_3 = 2K^3$ . This gives us the moments of the IUH of curve a of Figure 2 as

$$c = T/2 + K, \quad U_2 = K^2 + T^2/12,$$

$$U_3 = 2K^3, \text{ and}$$

$$m_2 = \frac{n^2 + 12}{3(n + 2)^2}, \quad m_3 = \frac{2}{(n/2 + 1)^3},$$

where  $n = T/K$ .

The relation between  $m_2$  and  $m_3$  has been evaluated and plotted as curve a on Figure 2.

The routed isosceles triangle (Fig. 2, Curve b; Fig. 4)—By the same reasoning as above we obtain  $c = T/2 + K$ ,  $U_2 = T^2/6 + K^2$ ,  $U_3 = 2K^3$ , and  $m_2 = (n^2 + 6)/6(n/2 + 1)^2$ ,  $m_3 =$

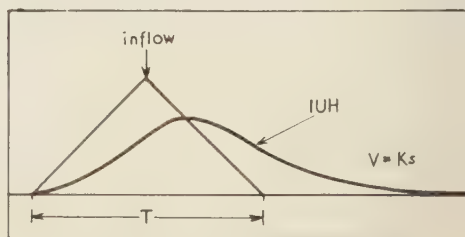


FIG. 4—The routed isosceles triangle

$2/(n/2 + 1)^3$ . The relation between  $m_2$  and  $m_3$  is shown as curve b on Figure 2.

*Successive routings*—If an instantaneous inflow is routed through a series of  $n$  linear reservoirs, each characterized by  $V = Ks$ , the outflow takes the form

$$u = \frac{V}{K\Gamma(n)} e^{-t/k} (t/K)^{n-1}.$$

The author has suggested elsewhere [Nash, 1958], the possibility of using this equation for the IUH. By repeated application of (1) and (5) we get  $c = nK$ ,  $U_2 = nK^2$ ,  $U_3 = 2nK^3$ ,  $m_2 = 1/n$ ,  $m_3 = 2/n^2 = 2m_2^2$ . The relation between  $m_2$  and  $m_3$  is shown as curve c on Figure 2.

*The equation of the log-normal frequency distribution*—The author has noticed that  $S$  curves (the integral of the IUH) are frequently nearly straight lines when plotted on log-normal probability paper. This implies an equation for the IUH of form

$$u = \frac{1}{t\sqrt{h\pi}} e^{-\frac{(\log t - g)^2}{h}}$$

where  $g$  and  $h$  are parameters. Chow [1955] gives the moments of this curve about the origin as

$$U'_n = \exp\left(\frac{hn^2}{4} + ng\right),$$

from which it can be shown that

$$c = e^{h/4} + g,$$

$$U_2 = e^{h+2g} - e^{h/2+2g},$$

and

$$U_3 = e^{9h/4+3g} - 3e^{h/4+g}(e^{h+2g} - e^{h/2+2g}) - e^{3h/4+3g}.$$

Whence  $m_2 = e^{h/2} - 1$ ,  $m_3 = e^{3h/2} - 3e^{h/2} + 2 = m_2^3 + 3m_2^2$ . The relation between  $m_2$  and  $m_3$  is shown as curve d on Figure 2.

For the routed rectangle and triangle there are two possible values of  $m_2$  for each value of  $m_3$ . This means that in fitting these forms to an actual IUH by equating first and second moments there are two possible values of  $T/K$  each of which gives exactly the same second moments. This is not really a disadvantage, as the position of the point corresponding to the actual IUH on Figure 2 shows which value of  $m_3$  (and therefore of  $T/K$ ) gives the best fit.

#### REFERENCES

- BERNARD, MERRILL, An approach to determinate stream flow, *Trans. Amer. Soc. Civ. Eng.*, **100**, 347-395, 1935.
- CHOW, VEN TE, On the determination of frequency factor in log-probability plotting, *Trans. Amer. Geophys. Union*, **6**, 481-486, 1955.
- COMMONS, G. G., Flood hydrographs, *Civ. Eng.*, **12**, 571-572, 1942.
- EZEKIEL, MORDECAI, *Methods of correlation analysis*, 2nd ed., John Wiley and Sons, 531 pp., 1941.
- MCCARTHY, G. T., Determination of unit hydrographs without flow records, appendix 3. *Engineering construction, flood control*, Engineer School, Fort Belvoir, Virginia, pp. 241-247, 1940.
- NASH, J. E., The form of the instantaneous unit hydrograph, *C. R. et Rapports, Assn. Internat. Hydrol. IUGG*, Toronto, 1957, Gentbrugge, **3**, 114-121, 1958.
- O'KELLY, J. J., The employment of unit hydrographs to determine the flows of Irish arterial drainage channels, *Proc. Inst. Civ. Eng.*, **4**, 365-445, 1955.
- SNYDER, FRANKLIN F., Synthetic unit-graphs, *Trans. Amer. Geophys. Union*, **19**, 447-454, 1938.
- TAYLOR, ARNOLD B., AND HARRY E. SCHWARTZ, Unit hydrograph lag and peak flow related to basin characteristics, *Trans. Amer. Geophys. Union*, **33**, 235-246, 1953.

(Manuscript received April 4, 1958.)





## On the Origin of Rock Magma

ROBERT J. UFFEN

*Department of Geophysics,  
The University of Western Ontario,  
London, Canada*

**Abstract** The remarkable correlation between the locations of ocean trenches, eugeosynclines, gravity anomalies, volcanic islands, and earthquake foci suggests a common origin. These phenomena are frequently interpreted as being due to compressive failure of the outermost crust of the earth. Yoder's [1952] hypothesis has been extended to include remelting upon release of pressure during compressive failure rather than tensile failure. The range of mean pressure, and the corresponding melting ranges for likely constituents of the crust and mantle, have been estimated. This hypothesis of magma generation is quantitatively feasible, but the data, especially the actual temperature distribution, are at present inadequately known.

**Introduction**—One of the major problems of geological science is the determination of the origin of magma, the formation of volcanoes, and the sources of the mobile rocks and fluid emanations which have resulted in the emplacement of igneous rocks in the earth's crust. The physical-chemical problem is that of specifying, preferably quantitatively, the melting temperatures for rocks of complex chemical composition for a region of pressure beyond that easily attainable in laboratory experiments. Consequently the method of attacking the problem usually involves the estimation of reasonable upper and lower limits for the physical properties through the extrapolation of the available low pressure laboratory data in accordance with the available theories of the solid and liquid states.

Earthquake seismology provides us with significant facts about the state of stress in the crust and mantle. The depth to the Mohorovicic elastic discontinuity, marking the base of the crust, is as little as 5 km under the oceans, about 30 km under continents, and as great as 60 km under great mountain ranges. The granitic layer is missing in oceanic areas [Ewing and Press 1955]. For short period displacements the crust and mantle are quite rigid, and molten chambers much greater than the wave lengths of seismic waves are unlikely to have escaped detection (about 600 m for a representative velocity of 6000 m/sec and a frequency of 10 cy/sec).

The remarkable correlation between the locations of earthquake foci, volcanoes, ocean trenches, mountain ranges, and gravitational anomalies is well known [Gutenberg and Richter

1950]. These features occur in a series of arcs with the ocean trenches and negative Bouguer gravity anomalies on the convex or oceanic side; then the volcanoes, positive gravity anomalies, and shallow earthquakes; and finally the deep focus earthquakes and shallow seas on the concave or continental side. A generalized section can be found in Wilson [1954]. The locus of earthquake activity dips toward the continent at an angle of about  $34^\circ$  to a depth of about 300 km and then more steeply, about  $60^\circ$  to a depth of 700 km [Benioff 1955]. In these marginal regions the crust of the earth has been under compressive stress, and the result has been the development of great thicknesses of sediments in geosynclines and folded mountain belts.

These earthquakes are evidence of relief of stress, rapid on the geological time scale, and their close correlation with the loci of volcanoes strongly suggests a common origin.

**The stress-relief hypothesis**—A hypothesis has been presented which may explain the above-mentioned correlation which, while it seems to have been presented in a qualitative way earlier in the geological literature, is here attributed to Yoder [1952] who appears to have made the first attempt at a quantitative study. Yoder proposed that the materials of the earth's crust have been locally re-melted by a drop in melting temperature, below the ambient temperature, upon release of pressure. A volume of material at depth might possess a temperature in excess of that required for melting at a reduced pressure, so that upon release of the pressure melting takes place.

This hypothesis has received insufficient

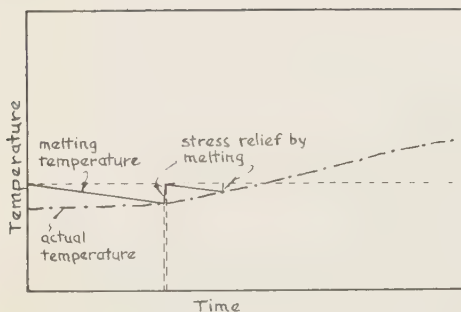


FIG. 1—The melting temperature with the superposition of tensile stress on the normal hydrostatic pressure

attention probably because Yoder considered only the effect on the reduction of the melting temperature of tensions, developing fractures which are accompanied by tectonic earthquakes. Figure 1 shows, qualitatively, how the melting point would decrease with time due to the superposition of a tensile stress on the normal hydrostatic pressure for anhydrous rocks. In this case, as soon as melting commenced the stress would be relieved, the melting temperature would rise, and the process would stop. It has been assumed for Figures 1 and 2 that the actual temperature increases with time. Possible mechanisms by which this might occur are suggested later.

In view of the correlation of volcanic activity with apparent compressive failure along continental margins, we must also inquire into the effects of compression on the melting temperatures. Figure 2 shows, qualitatively, how the melting point would increase with time due to the superposition of a compressive stress on the normal hydrostatic pressure for anhydrous rocks. The significant difference is that, upon relief of a tensile stress, the melting temperature rises to its normal value which may be above the ambient temperature, whereas upon relief of compressive stress the melting temperature decreases to its normal value which may be below the ambient temperature, thus permitting a period of remelting. In the first case continued melting can only occur if the ambient temperature rises until it exceeds the normal melting temperature. In the second case the ambient temperature could rise above the normal melting temperature without producing melting, unless there occurred a sudden

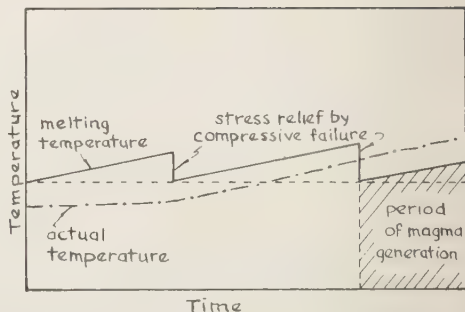


FIG. 2—The melting temperature with the superposition of compressive stress on the normal hydrostatic pressure

relief of compressive stress by yielding or fracturing.

*The mean pressure within the earth*—For many purposes it is a useful simplification to treat the stress within the earth as a simple pressure, that is, as if the crust behaved as a dense fluid with a specific gravity of between two and one half and three. Then the pressure at depth  $h$  is  $p = g\bar{\rho}h$ , where  $\bar{\rho}$  is the mean density to depth  $h$ , and  $g$  is the acceleration due to gravity. However, it is necessary here to take into account the deviations from fluid behavior. Birch [1955] has shown how the mean pressure can be estimated, given a knowledge of the shear strength at any depth.

Any stress can be resolved into the mean pressure, producing negative dilation, plus a system of shearing stresses, producing distortion. Consider a point in the earth where the vertical component of stress equals the weight of the overlying layer  $\bar{\rho}gh$  and is along a principal stress axis. The other two principal stress axes would be horizontal. If  $\sigma_1 > \sigma_2 > \sigma_3$  are the principal stresses then [Jaeger, 1956, p. 17]

$$\tau_1 = \frac{1}{2}(\sigma_2 - \sigma_3),$$

$$\tau_2 = \frac{1}{2}(\sigma_1 - \sigma_3),$$

and

$$\tau_3 = \frac{1}{2}(\sigma_1 - \sigma_2),$$

where  $\tau_1$ ,  $\tau_2$  and  $\tau_3$  are the principal shear stresses. If we set  $\sigma_3 = \bar{\rho}gh$ ,  $\sigma_2 = \bar{\rho}gh \pm 2S = \sigma_1$  as in Figure 3 then

$$\tau_1 = \pm S = \tau_3$$

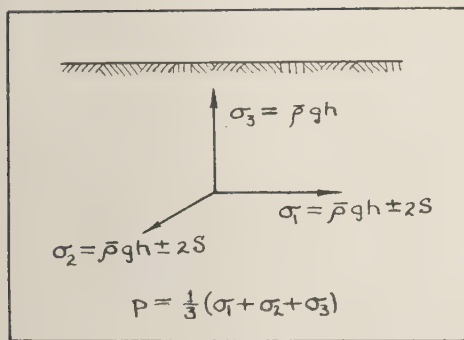


FIG. 3—The principal stress axes

and

$$\tau_s = 0.$$

If  $S$  is the maximum shear stress the material can withstand, then the mean pressure

$$p = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3)$$

must lie within the limits  $\rho gh - \frac{4}{3}S < p < \rho gh + \frac{4}{3}S$ . If we have a suitable value for the maximum shearing stress, the limits for the mean pressure at various depths can be evaluated.

Birch took  $S = 3000$  bars as a reasonable value for granitic rocks under moderate pressure; however, we should note that little is known for rocks about the variation of  $S$  with pressure. It probably increases with depth in the earth, passes through a maximum, and then, because of increase in temperature, decreases at great depth. At depths in excess of 700 km, the lack of earthquake foci suggests that the material is quite ductile and never fails by rupture.

As a demonstration of the type of behavior which might occur, Figure 4 shows the stress-strain curves for the failure of limestone under various mean pressures [Robertson, 1952]. At low mean pressures the limestone behaved as a brittle material with failure by rupture. At high mean pressure the behavior was ductile with considerable deformation before rupture occurred but with greatly increased strength. At extremely high mean pressures it is doubtful that failure by rupture ever occurs. For short time experiments it has been found (D. Griggs and J. Handin, personal communication, May, 1958.) that basalt may have a strength as great as 10,000 bars under a confining pressure of 5000 bars at a temperature of 500°C.

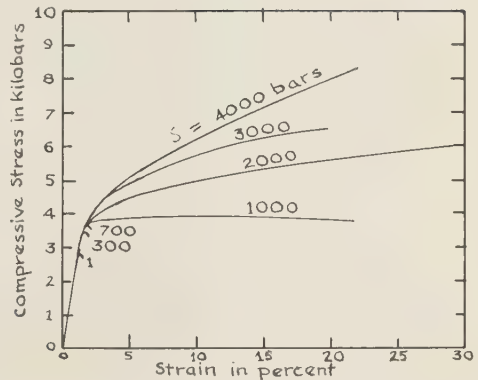


FIG. 4—Effect of confining pressure on the behavior of Solenhofen limestone under compression (after E. Robertson)

Table 1 gives the hydrostatic pressure and limits for the mean pressure at various depths in the earth, using in Case A,  $S = 3000$  bars, which is likely to be an underestimate, and for Case B, the arbitrary expression,  $S = (3000 + 100h)$  bars.

TABLE 1

Depth $h$ km	Mean pressure in kilobars, $\rho gh \pm \frac{4}{3}S$	
	Case A $S = 3.0$ kb	Case B $S = (3.0 + 0.1h)$ kb
0	0	0
10	$2.7 \pm 4.0$	$2.7 \pm 5.3$
20	$5.5 \pm 4.0$	$5.5 \pm 6.7$
30	$8.4 \pm 4.0$	$8.4 \pm 8.0$
40	$11.5 \pm 4.0$	$11.5 \pm 9.3$
50	$15.0 \pm 4.0$	$15.0 \pm 10.7$
60	$18.0 \pm 4.0$	$18.0 \pm 12.0$
80	$25.0 \pm 4.0$	$25.0 \pm 14.7$
100	$31.0 \pm 4.0$	$31.0 \pm 17.3$
200	$65.0 \pm 4.0$	$65.0 \pm 30.6$
400	$141.0 \pm 4.0$	$141.0 \pm 57.4$

We see that for Case A, at a depth of 30 km, corresponding to the bottom of the continental crust, the mean pressure could be as high as 12,400 bars just before failure (during mountain building, for example) and could drop to 8400 bars upon failure. If subsequent tension were superimposed, the mean pressure could drop as low as 4400 bars. For Case B the mean pressures would be 16,400 bars, 8400 bars, and 400 bars respectively.



*Effect of mean pressure on melting temperature*—The melting temperature of most substances increases with pressure. The change in melting point with pressure is given by the Clausius-Clapeyron equation

$$\frac{dT_m}{dp} = \frac{\Delta V}{\Delta S},$$

where  $\Delta V$  is the volume of the liquid minus that of the solid and  $\Delta S$  is the corresponding difference in entropy. This ratio has been determined, at ordinary pressures, for a few important minerals so that we are able to use these values to obtain melting temperatures at high pressures for anhydrous rocks by extrapolating from known melting temperatures at ordinary pressures. For pressure  $p$  in bars we have the data shown in Table 2. (See *Daly*, 1933, and *Verhoogen*, 1954.)

TABLE 2

Mineral or rock	$T_m$ °C	$\frac{dT_m}{dp}$ °C per bar	Source
Diopside ( $\text{CaMgSi}_2\text{O}_6$ )	1891.5	.013	<i>Yoder</i> , 1952
Forsterite ( $\text{Mg}_2\text{SiO}_4$ )	1890	.0047	<i>Bowen</i> , 1935
Basalt	1000	.01	<i>Yoder</i> , 1952
Granite (containing 7% water)	600	.02(?)	<i>Daly</i> , 1933, p. 68
"Silicates"	1300	.01	<i>Jeffries</i> , 1929

It is probable that  $dT_m/dp$  decreases with depth in the earth, that is, with increasing pressure [*Uffen*, 1952], but this is unlikely to be important for depths less than 700 km. We see from Tables 1 and 2, that for Case A, at the base of the continental Crust (a depth of about 30 km), basaltic material could have a melting temperature as high as 1120°C, dropping to 1080°C upon compressive failure, and even as low as 1040°C if tension were superimposed on the hydrostatic pressure. For Case B, the corresponding melting temperatures would be 1160°C, 1080°C, and 1000°C respectively.

*The actual temperature in the earth*—Unfortunately, our knowledge of the actual temperature is quite inadequate. Two methods which have been used to make estimates are the extrapolation of temperatures measured in mines and bore-holes at the surface of the earth and the computation of theoretical temperatures for various earth

models. In Figure 5 are shown two examples of many recent investigations [*Misener* and others, 1951] and [*Jacobs*, 1954, Model II]. The deviation between the linear extrapolation of the observations and the theoretical curves is a fair indication of the present state of ignorance. It is highly probable that the temperature gradient decreases with depth, but to what extent is still a matter of conjecture. The calculated curve was for a cooling earth model in which the crust was taken as 60 km thick and progressively less acid and less radioactive with depth. The actual temperature probably lies somewhere between these two and may be quite different for oceanic and continental areas. One such estimate is shown in Figure 5.

*The formation of magma*—Re-melting, upon release of pressure during compressive failure, will only occur if the actual temperature rises during compression until it exceeds the melting temperature for the unstressed state (see Fig. 2). The entire argument depends on this assumed rise in temperature with time. However, the particular mechanism by which this occurs is not crucial for the validity of the hypothesis, and there are at least four processes by which the actual temperature of rock, destined to become magma, could have risen the required amount: (1) by depression of crustal material to depths where the temperature is higher; (2) by adiabatic self-compression in the earth's gravitational field; (3) by slow transport of hot solid material from the base of the mantle; or (4) by heating from radioactive disintegration. Any one or a combination of these or other mechanisms may in fact occur. Of course, it is possible that this heating mechanism could raise the temperature beyond the melting point of even the stressed state, in which case melting would occur, but without the compressive failure and accompanying earthquakes.

Figure 5 shows for Case A the approximate melting ranges of three important rock compositions: granitic, basaltic, and dunitic. The existence of rock of any particular composition below the depth where its maximum melting point crosses the actual temperature curve would be inconsistent with the seismic evidence unless it formed a 'honey-comb' of fluid in the otherwise solid mantle. In general, such a fluid portion would be less basic and less dense than the material from which it was generated. The depths

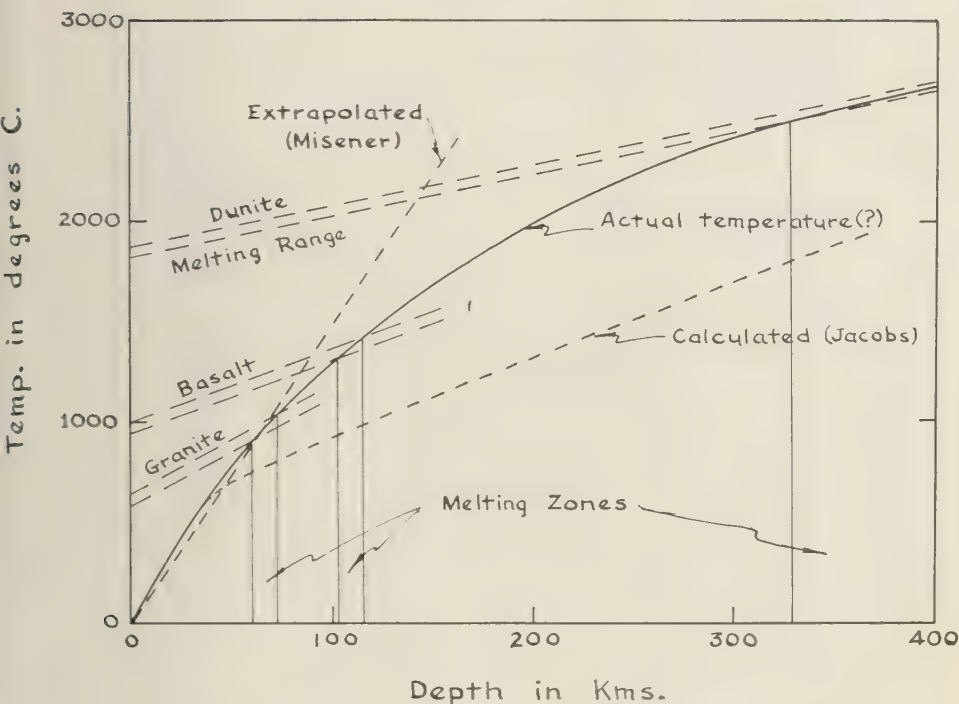


FIG. 5—Zones of potential remelting within the earth

within the earth at which the melting ranges intersect the actual temperature curve are the depths at which the above three types of magma could have earthquakes associated with their formation. Regardless of the choice of the actual temperature curve, it is clear that the dunitic magma could be generated at great depths. Basaltic magma could be generated at intermediate or greater depths. Granitic magma could be generated at shallow, intermediate, or great depth. The suggested temperature curve, which is shown intermediate to the others, would be consistent with the formation of granitic magma near the base of the continental crust, basaltic magma from about 100 km to 120 km, and dunitic magma over a large range of depth between 300 and 600 km where the earthquake activity ceases.

Yoder [1952] considered the problem of the source of the latent heat of melting. Upon decrease of mean pressure, the zone for which the actual temperature exceeded the melting point would possess heat available for melting equal to the product of the excess temperature and the

specific heat of the rock. For example, basaltic material at a depth of about 100 km could have an excess temperature of 50°C and, as the specific heat of the hot rock would be about 0.30, there would be about 15 cal/g available to melt a portion of the rock. If the latent heat of melting is 100 cal/g [Daly, 1933, p. 64], approximately 15 pct of the rock could be melted. The remainder of the rock would have undergone considerable ductile deformation.

*The rise of magma*—Upon melting, the magma would tend to rise in the earth's gravitational field because of the decrease in density relative to the surrounding rock. Depending on the chemical composition of the crust and mantle, the state of stress, and the actual temperature, there would exist at some depth a zone of potentially less dense rock available to rise when sudden failure provided the melting conditions and a zone of weakness. Presumably such a molten magma would rise until it reached a level where its density deficiency relative to the surrounding rock was negligible, or where the actual temperature was below the melting point of the magma.

Two competing processes would determine whether the magma solidified before reaching the surface: the rate of rise of the light magma in the gravitational field and the rate of loss of heat from the magma to its relatively cool surroundings.

Should the magma reach the surface in large volume, in sufficient time the original crust could find itself sinking locally as in a tectogene or even on a global scale. If we assume the radius of the earth to have remained essentially constant, then the original crust would be in a region of higher temperature and confined to a smaller circumference, that is, under compressive stress. The net result would be a cyclic overturn of material with the earth's gravitational field the source of energy, and the eventual result would be the differentiation of the low-melting, acid fraction toward the surface.

*Conclusions*—No attempt has been made to specify the exact nature of the magma which would be generated at any depth. It is recognized that it has been an oversimplification to treat a rock as though it had a single melting temperature at a given pressure. All the difficulties pointed out by Verhoogen [1954] concerning the possible reversals of melting relations at depths of about 200 km remain unresolved. It seems clear, however, that the first liquid to form by partial fusion of the mantle above 200 km would be either basaltic or basaltic enriched in alkalis and alumina.

In order to account for the generation of granitic magma at depths less than 60 km, it would be necessary to admit either a steeper temperature gradient near the earth's surface or a lower melting temperature for granitic material, say by virtue of increased water content. The latter possibility has been stressed by Tuttle and Wyllie [1957]. According to the stress-relief hypothesis it should be relatively commonplace to find igneous granite emplaced in highly deformed pre-existing rocks from which it was generated, while the occurrence of dunitic rocks would be expected to be rare in view of its great depth of origin.

The stress-relief hypothesis is qualitatively capable of explaining the observed correlation between the occurrence of volcanoes and earthquakes. Quantitatively it appears feasible, although the experimental data are insufficient for adequate precision. The lack of adequate

knowledge of the actual temperature is the greatest stumbling block. If, however, the hypothesis is true, it might be possible to reverse the problem and seek information about the depth of origin of various magmas from other sources, thus eliminating earth models whose computed temperature distributions for the present age of the earth do not agree.

*Acknowledgments*—I wish to express my appreciation to C. M. Carmichael, for many stimulating discussions.

# REFERENCES

- BENIOFF, H., Seismic evidence for crustal structure and tectonic activity, *G. S. A. SP No. 62*, 1955.
- BIRCH, F., Physics of the crust, *G. S. A. SP No. 62*, 1955.
- BOWEN, N. L., AND J. F. SCHAIERER, The system,  $MgO-FeO-SiO_2$ , *Am. J. Sci.*, 29, 151, 1935.
- BULLEN, K. E., *Introduction to the theory of seismology*, Cambridge University Press, 1947.
- DALY, R. A., *Igneous rocks and the depths of the earth*, McGraw-Hill Book Co., Inc., New York 1933.
- EWING, M., AND F. PRESS, Geophysical contrasts between continents and ocean basins, *G. S. A. SP No. 62*, 1955.
- GUTENBERG, B., AND C. F. RICHTER, *Seismicity of the earth*, Princeton University Press, 1950.
- JACOBS, J. A., AND D. W. ALLAN, Temperature and heat flow within the earth, *Trans. Roy. Soc. Canada*, Vol. XLVIII, 1954.
- JAEGER, J. C., *Elasticity, fracture and flow with engineering and geological applications*, Methuen & Co., Ltd., London, 1956.
- JEFFREYS, H., *The earth*, 2nd edition, Cambridge University Press, 1929.
- MISENER, A. D., L. G. D. THOMPSON, AND R. J. UFFEN, Terrestrial heat flow in Ontario and Quebec, *Trans. Am. Geophys. Union*, 35, No. 5, 1951.
- ROBERTSON, E., An experimental study of flow and fracture in rocks, Ph.D. Thesis, Harvard University, 1952.
- TUTTLE, O. F., AND P. J. WYLLIE, Hydrothermal studies in the systems  $NaAlSi_3O_8$  (albite)  $H_2O-HF$  and granite- $H_2O-HF$  (abstract), *Bull. Geol. Soc. Am.*, 68, No. 12, 1957.
- UFFEN, R. J., The thermal state of earth's interior, Ph.D. Thesis, University of Western Ontario, 1952.
- VERHOOGEN, J., Petrological evidence on temperature distribution in the mantle of the earth, *Trans. Am. Geophys. Union*, 35, No. 1, 1953.
- WILSON, J. TUZO, in *The earth as a planet*, Chapter 4, University of Chicago Press, 1954.
- YODER, H. S., Change of melting point of diopside with pressure, *J. of Geol.*, 60, 364-374, 1952.

(Manuscript received August 11, 1958, and as revised, October 6, 1958.)

# The Calorimetry of Steaming Ground in Thermal Areas

R. F. BENSEMAN

*Department of Scientific and Industrial Research  
Dominion Physical Laboratory,  
Private Bag, Lower Hutt, New Zealand*

**Abstract**—A portable calorimeter that measures heat output from the steaming ground found in areas of natural thermal activity is described. With a minimum of disturbance to the site, heat flow is measured in a range from  $10$  to  $70 \times 10^{-3}$  cal/cm<sup>2</sup> sec with an accuracy better than 10 per cent. Below this range, the accuracy decreases. The relationship of heat output to the soil temperatures measured at a depth of 35 cm is shown for measurements at 27 different sites.

**Introduction**—The total natural output of heat from a thermal region can be a useful guide to its potentialities as a source of underground steam for the generation of electric power. The main thermal regions of New Zealand occur in country formed of very permeable pumice breccia, and they are characterized by extensive tracts of ground through which steam slowly percolates. The amount of heat lost in this way is comparable to the heat output of the more obvious features such as geysers and hot springs. The present apparatus is considered to be the best of a number of calorimeters that have been used in New Zealand to measure the heat output from steaming ground. Important features are (1) it does not require the ground surface to be smoothed or leveled, and it does not produce a partial vacuum over the site; therefore there is good reason to suppose that the flow of steam is unaffected by the presence of the calorimeter, and (2) being light and portable, the instrument is suited to use in rugged and bush-covered country.

**Description of apparatus**—The calorimeter is essentially a box, formed of sheet aluminum on a light frame, and having an open bottom 25 by 80 cm. The box is pressed slightly into the ground so as to seal off the portion of ground below it. Ports in the top allow a fan to draw air through the box at a measured rate, and the temperature and humidity of the air are measured as the air enters and leaves the box. From these measurements the flow of heat into the box can be calculated.

Air enters at *A* (Fig. 1) past differential thermocouples which detect the wet-bulb and dry-bulb temperatures. It passes a hot-wire anemometer and then traverses the chamber *S*

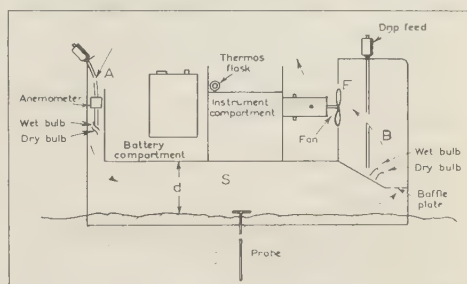


FIG. 1—Schematic section of calorimeter box

where it gains heat from the ground. It is exhausted from the system by the fan *F* after passing a second set of thermocouples at *B*. To reduce heat transfer through the sides and top of the chamber *S*, these are lined with quarter-inch cork sheet. The fan and the anemometer heater are powered by the battery shown, and the instrument compartment includes a multi-position switch that selects the circuit whose emf is to be measured. Figure 2 is a photograph of the calorimeter box and the portable potentiometer used for measuring the emfs.

**Design details**—Previous calorimeters had shown that steaming ground may emit as much as  $50 \times 10^{-3}$  cal/cm<sup>2</sup> sec, and it was estimated that the present calorimeter would need to pass seven liters of air per second if the dry-bulb temperature were not to rise by more than 20°F in the most unfavorable weather conditions (85°F and 80 pct relative humidity). Using entry and exit ports of area 100 cm<sup>2</sup> it was found that this rate of flow produced no measurable pressure deficiency in the sampling chamber. The chamber was made 12 cm deep so that the air velocity would not exceed one mile an



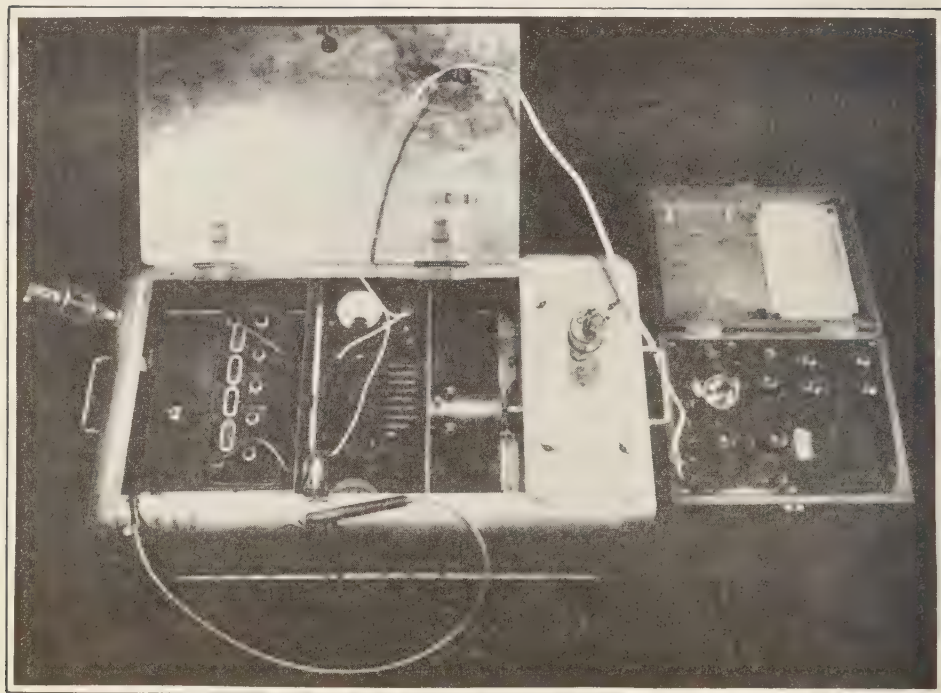


FIG. 2—Completed box with its potentiometer

hour, comparable to the wind speed on a calm day.

The hot-wire anemometer has been described by Benseman and Hart [1955]. Briefly it consists of a short length of one inch square-section tubing in which is an electrically heated grid flanked by a differential thermopile. As the air speed through the tube varies, so does the emf generated by the thermopile. This anemometer has proved superior to more conventional types since it has no moving parts, retains its calibration after rough handling, and is most sensitive at very low rates of air flow. It was calibrated *in situ* by drawing air through the box at known rates, the fan and motor having been removed.

The wet-bulb and dry-bulb temperatures at inlet and outlet were each measured by five copper-constantan thermojunctions spaced across the air stream and held rigidly on suitable brackets. The wet junctions were provided with muslin envelopes kept moist by drip attachments fixed to the body of the box. In the present apparatus the air speed does not exceed 3 ft/sec,

so humidity is deduced from tables used by the Meteorological Office for wet- and dry-bulb temperatures observed within a screen.

The reference junctions of all the thermocouples are embedded, together with a mercury-in-glass thermometer, in paraffin wax within a small vacuum flask housed in the top of the box. All temperatures can therefore be determined absolutely, but in practice it is convenient to read (1) the temperature of the mercury thermometer, (2) the emf of the dry thermopile at the inlet, (3) the differential emf between dry thermopiles at inlet and outlet, and (4) the differential emf between wet and dry thermopiles at the inlet and at the outlet.

*Laboratory tests*—Tests in which moisture and heat were presented to the calorimeter in a variety of proportions showed that it would account for heat flows in the range  $10$  to  $70 \times 10^{-3}$  cal/cm<sup>2</sup> sec with an accuracy better than ten per cent. Below  $10 \times 10^{-3}$  cal/cm<sup>2</sup> sec the accuracy was  $\pm 1 \times 10^{-3}$  cal/cm<sup>2</sup> sec.

*Use in the field*—An area of steaming ground was chosen which could be reached only by boat and was therefore unlikely to be disturbed by casual visitors. Fourteen sites were pegged at ten-meter intervals in a straight line over ground that was not too active to support some stunted trees and bushes. Thirteen additional sites were chosen on a patch of very hot ground that grew only a little moss. These thirteen sites were surveyed on only one occasion, for it proved difficult to repeat individual measurements; the ground surface was soft and sticky, and merely walking over it seemed to affect the heat output at sites nearby. The 14 cooler sites were re-measured on a number of occasions spread over more than a year.

The temperature of the ground 35 cm below the surface was recorded also, for it was hoped that measurements at this depth might prove to be related to the heat flow. It is to be expected from theoretical arguments [Benseman, 1956] that the flow of steam through the ground should control the gradient near the surface.

The measurements are fully reported and discussed in a report that is available from this Laboratory on request [Benseman, 1958]. Here it is sufficient to summarize the conclusions.

*Operational notes and results of measurements*—The following summary comments are made:

(1) The calorimeter box is easily set in place, but it may take an hour before a steady reading can be obtained. During this time the fan must be in continuous action.

(2) Measurements of heat output made in the afternoon tend to be a little higher (10 to 15 pct) than measurements made in the morning. This extra heat is probably that fraction of the solar heat that the surface of the box fails to reflect.

(3) Winds can bring gusts of warm and cold air past the site of the measurements. The resulting fluctuations in the outputs of the thermocouples make work difficult except in calm weather or in light winds. Some thermal lagging of the thermocouples might reduce this difficulty.

(4) Heat outputs at the 14 sites showed large daily and monthly changes, half of the changes between readings exceeding 30 pct. These variations suggested a redistribution of steam flow in the ground rather than a significant change in total flow. No correlation was found

between the heat-flow variations and weather conditions, such as wind, cloud, rainfall, and barometric pressure or rate of change of pressure.

(5) The scatter diagram of Figure 3 shows

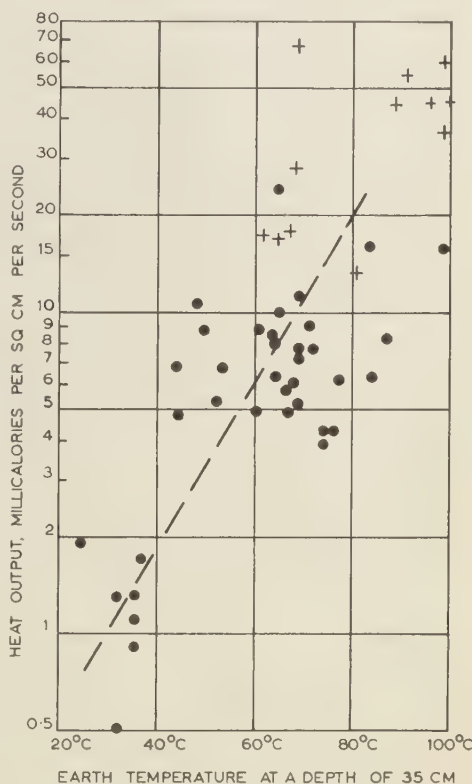


FIG. 3—A scatter diagram relating the heat output of each site with the temperature at 35 cm; crosses indicate sites in the hotter area, circles refer to sites in the cooler area.

the connection between the heat output at a site and the temperature at a depth of 35 cm below the surface. Earth temperatures can be measured much more quickly than the heat output, and in spite of the lack of exact correlation between these variables, the measurements of earth temperatures may be a rapid and sufficiently accurate means of estimating an overall heat flow. Figure 3 cannot of course be expected to apply to other thermal areas where the soil has a very different texture or permeability.

## REFERENCES

- BENSEMAN, R. F., AND H. R. HART, A thermocouple anemometer, *J. Sci. Instr.*, **32**, 1-17, 1955.
- BENSEMAN, R. F., The natural heat output of Orakei Korako, *Tech. Rep. R258*, Dominion Phys. Lab., Dep. Sci. Ind. Res., Wellington, New Zealand, pp. 11-12, 1956.
- BENSEMAN, R. F., Measurements of heat output from steaming ground, *Tech. Note 119*, Dominion Phys. Lab., Dep. Sci. Ind. Res., Wellington, New Zealand, pp. 1-17, 1958.

(Manuscript received August 18, 1958.)

# The Origin of Some Natural Carbon Dioxide Gases

WALTER B. LANG

*U. S. Geological Survey, Washington, D. C.*

**Abstract**—Natural carbon dioxide gas issues from vents and springs in sedimentary rocks. Exploratory wells drilled for oil have produced large volumes of carbon dioxide in subsurface. Where these gases are found in quantity, limestones and igneous rocks are in close association and the origin of the gases is thus suggested. Until the advent of the mass spectrometer there was no means to investigate this problem. Gas samples were taken from representative areas and the analyses showed that the gases were derived from limestone. The isotopic ratios did not give evidence of fractionation.

**Introduction**—The natural outflows of carbon dioxide in the United States are commonly associated with springs, but of late years the greatest source of information on the abundance and distribution of carbon dioxide in the sedimentary rocks has come as a consequence of prospecting for petroleum.

Two wildcat oil wells drilled in northeastern New Mexico in the middle twenties produced, under considerable pressure, gases rich in carbon dioxide. The observed geologic relations of limestone-bearing sediments and a prominent display of igneous flows in the area led to the natural assumption that the heat of igneous intrusion may have decomposed limestone with resulting formation of carbon dioxide gas [Lang, 1926].

Nier and Gulbranson [1939] showed that the  $C^{12}/C^{13}$  isotopic-abundance ratio for carbon of carbonates was different from that of the carbon of other substances. Subsequent researches of Baertschi [1951] and Craig [1953] further confirmed this observation for unmetamorphosed marine limestones.

Table 1 shows the abundance ratios of isotopes 12 and 13 for various sources of carbon, as determined by Nier and Gulbranson. From these results it is evident that limestone contains more of the heavier isotope and that the geologic age of the specimens is without effect on the isotopic abundance ratios.

The ratios of the additional analyses listed in Table 2 are slightly lower in value than those of Table 1. This does not reflect fundamental differences in these ratios but is the result of the use of a different spectrometer. A comparable gap exists between the ratios for limestone and the

organic carbons in both tables, a fact more clearly demonstrated in Table 2. There are now available in the literature several hundred isotopic analyses of limestones and carbonates. They all serve to support the fact that the carbon isotopic ratios for limestone are so grouped that they can be distinguished from those of the organic carbons.

A study of carbon dioxide gases from sediments involves two considerations: (1) discovery of the source of the gas, whether from limestone or some other material, and (2) the nature of the agency that most likely produced it. To promote this comparative study of the geologic environment with the identification of the gas, the writer collected samples of carbon dioxide from the more representative areas as opportunity permitted. The samples were sent to Nier for analysis, and the results given in Table 3 were obtained (Nier, personal communications, 1948–1954).

Nier expressed his original results as  $C^{12}/C^{13}$  ratios and used the gas formed from Solenhofen limestone as a standard reference. The trend has since been to express the results as  $C^{13}/C^{12}$  ratios. The  $C^{12}/C^{13}$  ratios may be converted to  $C^{13}/C^{12}$  ratios for comparison, or vice versa, by a conversion formula [Craig, 1953].

## GEOLOGIC DESCRIPTION

**Imperial Valley**—In the Imperial Valley of Southern California and contiguous Mexico, hot springs and mud pots with carbon dioxide gas issue from the alluviated valley floor. The bed rock in the Imperial Valley area is not known. The presence of bedded limestone or dolomite of possible Precambrian to Permian age in the mountains surrounding Imperial Valley [Mendenhall, 1909, 1910; Harder, 1912; Brown, 1923]



TABLE 1— $C^{12}/C^{13}$  ratio for various sources of carbon<sup>a</sup>

Sample	Source	Age <sup>b</sup>	Series tests <sup>c</sup>		
			1st	2nd	3rd
		<i>m y</i>			
<i>Igneous group</i>					
Graphite	Ceylon	500	89.7	89.9	....
Meteorite	Crosby's Creek, Tenn.	900	89.6	89.1	....
Diamond	Kimberly Mines, Africa	60	....	89.0	....
<i>Limestone group</i>					
Grenville	New York	1200	87.2	88.3	88.2
Ordovician	Vermont	380	88.6	88.5	....
Clam shell	Boston, Mass.	0	88.7	88.6	....
<i>Plant group</i>					
Anthracite coal	Tennessee	250	92.1	91.7	91.5
Pine wood	Massachusetts	0	91.8	91.2	....
<i>Unclassified group</i>					
Dry ice	Virginia coal	230	91.6	90.8	....
Petroleum	West Texas	200	....	91.2	....
Air	Mass., March 14, 1938	0	92.5	....	....
Air	Mass., March 22, 1938	0	....	89.8	89.9
Na <sub>2</sub> CO <sub>3</sub>	Michigan limestone	300	88.3	88.3	....
Clam flesh	Boston, Mass	0	....	90.1	....

<sup>a</sup> From Nier and Gulbranson [1939].<sup>b</sup> As estimated by A. C. Lane [Nier and Gulbranson, 1939].<sup>c</sup> Runs made on same samples.

(Davis, personal communication, 1957) suggests that such marine sediments probably occur at depth in the valley.

The high-temperature spring waters in the valley are indicative of igneous activity at depth although it is conceivable that faulting may have

TABLE 2— $C^{12}/C^{13}$  for carbon from various sources<sup>a</sup>

Source	Number of samples	Range of $C^{12}/C^{13}$ for group ratios	Average ratio for samples in group
Limestone	10	88.8–89.4	89.2
"Zeolitic calcite"	1	....	89.9
Marine shell	1	....	89.5
Sea water	1	....	89.3
Meteoritic carbon	7	89.8–92.0	91.3
Graphite	1	....	90.2
Coal (humic origin)	10	91.3–92.0	91.8
Wood	7	91.6–92.2	91.8
Petroleum	6	92.0–92.8	92.5
Bituminous shales	7	92.1–92.7	92.5
Torbanite and kerosene shales	3	91.3–92.0	91.7
Lycopodium spores	1	....	93.1
"Balkashite" algae	1	....	92.8
CO <sub>2</sub> in Minneapolis air	1	....	91.5

<sup>a</sup> From Murphey and Nier [1941].

TABLE 3—Isotopic analyses of natural carbon dioxide gas

Sample location	Year sampled	$C^{12}/C^{13}$ ratio
Imperial Valley, Calif.	1948	89.2
Upper gas, Ute Valley, N. Mex.	1949	89.1
Lower gas, Ute Valley, N. Mex.	1949	89.7
Tampico, Mexico	1950	89.6
Saratoga Springs, N. Y.	1950	89.5
Soda Springs, Idaho	1953	89.9

contributed a minor increment. Cardox-Western, Inc. (personal communication, 1950), who were engaged in the production of carbon dioxide from numerous wells drilled west of Niland, are of the opinion that the gas is now actively being generated. In drilling their wells, some of which have gone below 2000 ft in calcareous shales, they have found the gas present in varying amounts. This raises the question of whether the

gas may be seeping out at the surface from other than conspicuous vents.

*Ute Valley, New Mexico*—Carbon dioxide gas is produced in Ute Valley from two relatively flat-lying porous zones. The upper gas zone in sandstone beds of Triassic age lies at an average depth of 1000 ft. This gas is under a pressure of 50 lb/in<sup>2</sup>. The lower zone, which is in Permian sedimentary rocks, lies at a depth of 2000 ft and has a gas pressure of some 500 lb/in<sup>2</sup>, which is less than equivalent to a full hydrostatic head. These gas reservoirs are within the Sierra Grande structure which extends for more than 100 mi from southeastern Colorado, where carbon dioxide is present, to the vicinity of Wagon Mound and Bueyeros, New Mexico. Sierra Grande and Capulin Mountain, extinct volcanoes, stand near the middle of the structure. The geologic section contains many limestones [Miller, 1937]. Igneous intrusions, which have served as feeders for the numerous volcanic cones and lava flows [Collins, 1949] that mark the surface, have penetrated these limestones. The carbon dioxide gas found in the reservoirs is dry and often nearly pure.

The original gas pressures of the reservoirs is not known, but some impression of the reservoir capacity is indicated by the fact that one well blew through an open hole for nearly a year without diminution of gas pressure before it was capped.

*Tampico, Mexico*—In the broad Coastal Plain area of Mexico, which encircles Tampico to the west, there is abundant carbon dioxide gas in depth in the thick sections of limestones. Here there is a gas reservoir of some 10 by 20 mi in size that is perhaps the largest of its kind in the world. The great volumes of gas are held in fractures and openings of Jurassic and Cretaceous limestones which are overlain by Tertiary sediments. These rocks are intruded by a wide variety of igneous plugs, dikes, and sills, mainly basic, that are post-Cretaceous, Miocene, and later in age. The temperature of the brines associated with the gases and oil is above normal (probably the result of volcanic activity).

In a small corner of this area in the vicinity of Quebracha, 16 miles south and a little west of Tampico, the Cia. Mexicana de Hielo Seco, S. A., operates six wells which produce carbon dioxide. The volume of gas production is indicated in the following quotation from a letter

from this company (1951): "One well has been producing 10 million cubic feet per day since 1926 and is still producing at that rate. Another well nearby is producing and has produced 5 million cubic feet per day since 1925. There are four other wells which produced like quantities from 1915 until 1940 and are now pinched in. The operating gas pressures have not dropped during all this time; in fact, there are indications of increase in pressure. The well producing 10 million cubic feet per day has a well-head pressure of approximately 1050 lb/in<sup>2</sup>."

Muir [1936] reports that the discovery well had an initial production of 50 million cu ft of gas per day, and four other wells blew 150 million cu ft of gas before they were shut in.

A representative analysis of the Quebracha gas shows the following composition:

Methane	2.4 pct
Ethane	1.5
Propane plus	0.1
Carbon dioxide	96.0

Total	100.0 pct
-------	-----------

*Saratoga Springs, New York*—Carbon dioxide gas at Saratoga Springs issues with the water that comes from Cambrian and Ordovician limestone. Basaltic dikes occur northwest of Saratoga Springs, but these were mapped by Cushing and Ruedemann [1914] as Precambrian and are therefore older than the Cambrian and Ordovician limestones from which the gas and waters are reported to come. Broughton (personal communication, 1951) states that no surface or subsurface igneous rocks other than the Precambrian have been discovered since the finding of Stark's Knob, 12 mi east of Saratoga Springs. The Knob was at first believed to be a volcanic plug but subsequently was found to be an erosional remnant of an overthrust from the east.

Colony [1930] in a report on a reinvestigation of the geology about Saratoga Springs authorized by the Saratoga Springs Commission, concluded that explanations for observed dolomitization, mineralization, and dissolved components of the water are inadequate without resort to volcanism. He further states, "The processes are thought to have had some connection with the same expiring volcanism regarded as being responsible for some of the dissolved components in the mineral waters, and it is judged that the evidence for their operation helps to support the

theory of a deep-rooted source for the constituents in the mineral waters."

*Soda Springs, Idaho*—Soda Springs, or Beer Springs as they were better known to the early trappers and explorers, were visited and described by *Fremont* [1842] and *Hayden* [1871]. *Peale* [1877] made investigations of these springs, which occur frequently at various points in or near the Bear River. Hooper Spring, about a mile north of the town of Soda Springs, has a large flow of clear, sparkling, ebullient water. The water of this spring has a normal temperature of 53°F (1877), but others, such as Steamboat Springs, have temperatures as high as 88°F.

In this region there are abundant limestones. Exposed at the surface adjacent to the spring are Ordovician and Devonian limestones. A few miles across the valley to the east Carboniferous and Triassic limestones crop out. Limestone is the predominant rock of the region.

Volcanic activity is almost everywhere in evidence as exemplified by lava flows, basalt cones, hot springs, sulfur deposits, and similar manifestations of heat. Regarding a sulfur deposit a few miles east of Soda Springs, *Mansfield* [1927] states, "... southeastern Idaho contains no large deposits of gypsum, and the small amount of that mineral associated with the sulphur appears rather to have had a common origin with it than to have served as a source. The proximity of these deposits to volcanic centers suggests that a volcanic origin for the hydrogen sulphide and associated carbon dioxide is plausible."

The period of greatest volcanic activity in eastern Idaho and Yellowstone Park was during the Pleistocene with a possible continuation into the Recent. Carbon dioxide, hot springs, and other evidences of mineralization are most likely manifestations of this period of great volcanic activity.

*Discussion*—From the foregoing review of the geologic relations where samples of natural carbon dioxide gas were obtained and analyzed, certain facts are evident. The values for all analyses fall within the range of the  $C^{12}/C^{13}$  isotopic-abundance ratios representative of carbon from carbon dioxide gas derived from marine limestones. In most of these carbon dioxide gas localities there is evidence of marine limestones in the local geologic sections, and in others it may be inferred. Except for Saratoga

Springs, the areas where carbon dioxide gas samples were taken show evidence of associated igneous activity.

On the basis of the distribution of carbon dioxide gases, the United States may be divided into three parts: (1) that east of the 100th meridian, where very few showings of carbon dioxide gas of any significance have been found and where no active igneous processes are obvious; (2) the Rocky Mountain and Basin and Range provinces, where there are numerous large reservoirs of carbon dioxide and where igneous activity is, for the present at least, dormant; and (3) the Pacific Coast, where carbon dioxide is now associated with active igneous processes. The geological and physical conditions under which these gases were found are manifold; they are contained in large and extensive reservoirs under different temperatures, pressures, in rocks of different kinds and ages, yet the isotopic analyses are similar and all abundance ratios are within the isotopic range of limestone. When the gases were formed and how long they have been in storage can only be inferred from the information now available. There is good reason to believe that some carbon dioxide gases have been imprisoned for millions of years (as have the hydrocarbon gases of petroleum, which may date back to the early Paleozoic). The carbon dioxide gases were confined to intimate association with other gases, hydrocarbons, brines, and rocks, including limestone, sandstone, and other reservoir rocks.

It is believed that the field and isotopic evidence indicate that the carbon dioxide gases originated from the breakdown of limestone in contact with or adjacent to igneous intrusions. It should, however, be realized that carbon dioxide derived from other sources, if held for a sufficient time within a limestone reservoir, might acquire by fractionation sufficient heavy isotopes to convert the abundance ratio to one representative of a limestone origin. This may happen, but the results obtained from analyses of these gases do not indicate that this has happened.

The postulated generation of carbon dioxide from limestone raises many questions for which the geologic record lacks adequate answers. The generation of such large volumes of carbon dioxide as those at Quebracha, Mexico, or elsewhere, would require the dissociation of an enormous quantity of calcium carbonate. This



poses the question of what becomes of the remaining calcium oxide. Calcium oxide combines with silica to form wollastonite, a relatively common contact metamorphic mineral. Stopping or absorption by magma is another process of elimination, but there are places where the regional geology does not permit this explanation.

## REFERENCES

- BAERSCHI, P., Relative abundances of oxygen and carbon isotopes in carbonate rocks, *Nature*, **168**, 288-289, 1951.
- BROWN, J. S., The Salton Sea region, California; a geographic, geologic, and hydrologic reconnaissance, with a guide to desert watering places, *U. S. Geol. Surv. Water Supply Paper 497*, pp. 41-58, 1923.
- COLLINS, R. F., Volcanic rocks of northeastern New Mexico, *Geol. Soc. Amer. Bul.*, **60**, 1017-1040, 1949.
- COLONY, R. J., Restudy of the geology of the Saratoga area and the problem of the mineral waters, *Rep. Saratoga Springs Commission, Legislative Doc.*, **70**, pp. 73-216, 1930.
- CRAIG, HARMON, The geochemistry of the stable carbon isotopes, *Geochim. Cosmochim. Acta*, **3**, 53-92, 1953.
- CUSHING, H. P., AND R. RUEDEMANN, Geology of Saratoga Springs and vicinity, *New York Museum Bul.* **169**, 177 pp. 1914.
- FREMONT, J. C., *Report of the exploring expedition to the Rocky Mountains*, House Doc. 166, pp. 134-139, 1842 (1845).
- HARDER, E. C., Iron-ore deposits of the Eagle Mountains, California, *U. S. Geol. Surv. Bul.* **503**, pp. 30-41, 1912.
- HAYDEN, F. V., *Preliminary report of the U. S. Geological Survey of Montana and portions of adjacent territories*, Hayden 5th an. rep., pt. 1, pp. 150-159, 1871 (1872).
- LANG, W. B., Unusual natural gases, *Bul. Amer. Assn. Pet. Geol.*, **10**, 1176-1177, 1926.
- MANSFIELD, G. R., Geography, geology, and mineral resources of part of southeastern Idaho, *U. S. Geol. Surv. Prof. Pap.* **152**, 409 pp., 1927.
- MENDENHALL, W. C., Ground water of the Indio region, California, *U. S. Geol. Surv. Water Supply Pap.* **225**, 56 pp., 1909.
- MENDENHALL, W. C., Notes on the geology of Carrizo Mountain and vicinity, San Diego County, California, *J. Geol.*, **18**, 336-355, 1910.
- MILLER, J. C., Carbon dioxide accumulations in geologic structures, *Amer. Inst. Min. Metal. Eng. Tech. Pub.* **841**, 28, pp., 1937.
- MUIR, J. M., *Geology of the Tampico region, Mexico*, Amer. Assn. Pet. Geol., 280 pp., 1936.
- MURPHEY, B. F., AND A. O. NIER, Variations in the relative abundance of the carbon isotopes, *Phys. Rev.*, 2nd series, **59**, 771-772, 1941.
- NIER, A. O., AND F. A. GULBRANSON, Variations in the relative abundance of carbon isotopes, *J. Amer. Chem. Soc.*, **61**, 697-698, 1939.
- PEALE, A. C., *Report on the geology of the Green River district*, Hayden 11th an. rep., pp. 509-646, 1877 (1879).

(Manuscript received July 29, 1958.)



## Letter to the Editor

## DISCUSSION OF "UPPER AND LOWER FREQUENCY CURVES FOR PEAK RATES OF RUNOFF"

BY W. D. POTTER

In a recent paper Potter [1958] offers an explanation of nonlinearity in a flood-probability graph described by [Rowe and others, 1957] as a "typical bilinear or dog-leg graph on probability paper." Further, the author derives a relation between what he refers to as "upper and lower frequency curves." Both papers imply that, provided the data comes from a homogeneous population, a linear relation is to be expected on the probability paper used, and departures from linearity are ascribed to lack of homogeneity. Admittedly, nonhomogeneity could be a cause, but there are others. Various studies [Alexander, 1957] have shown that sampling errors alone can give rise to 'dog-legs,' irrespective of the distribution, but the author apparently believes that the number of dog-legs obtained is too great to be ascribed to chance.

The statistics of flood hydrology is still empirical, and consequently the choice of distribution upon which the probability paper is based is primarily governed by its facility to produce linearity. Gumbel [1941] gave reasons for an *a priori* preference for the Fisher-Tippett Type I distribution, and the paper used by the author conforms to this type. However, we must examine more closely the compatibility between the hydrologic and statistical models for floods; if this distribution is inapplicable we are justified in seeking another.

There is considerable statistical evidence that the logarithmic transformation is applicable to flood data, possibly more so than to storm rainfall; consequently the log-normal and Fréchet (Fisher-Tippett Type II) distributions [Morlat and others, 1956] have found favor.

It seems premature to attempt to assign floods or meteorological elements to one of the three Fisher-Tippett types on a *posteriori* grounds as suggested by Jenkinson [1955] and Thom [1954]. For such purposes we would require a large number of independent samples and a knowledge of the behavior of the variate over a wide range. So long as selection is based on a *posteriori* grounds, linearity is the criterion; hence 'dog-legs,' if not due to sampling errors, are best regarded as indicating the unsuitability of the distribution for the data plotted.

## REFERENCES

- ALEXANDER, G. N., Flood flow estimation, probability and the return period, *J. Inst. Engrs., Australia*, **29**, 263-278, 1957.  
GUMBEL, E. J., The return period of flood flows, *Ann. Math. Statistics*, **12**, 163-190, 1941.  
JENKINSON, A. F., The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. Roy. Meteorol. Soc.*, **81**, 158-171, 1955.  
MM. MORLAT, BILLET, AND BERNIER, Les Crues de la Haute Durance et la theorie statistique des valeurs extremes, *Symposia D'Arcy*, **3**, 99-114, 1956.  
POTTER, W. D., Upper and lower frequency curves for peak rates of runoff, *Trans. Am. Geophys. Union*, **39**, 100-105, 1958.  
ROWE, R. R., G. L. LONG, AND T. C. ROYCE, Flood frequency by regional synthesis, *Trans. Am. Geophys. Union*, **38**, 879-884, 1957.  
THOM, H., Frequency of maximum wind speeds, *Proc. Am. Soc. Civil Engrs.*, **80**, Sep. No. 539, 1954.

G. N. ALEXANDER

"Forsythia", Bayview Road,  
Balgrave, Victoria Australia (June 2, 1958)



# Contents

	PAGE
Densities and Temperatures of the Upper Atmosphere Inferred from Satellite Observations,..... <i>G. F. Schilling and T. E. Sterne</i>	1
The Diurnal and Annual Variations of foF2 over the Polar Regions, <i>S. C. Coroniti and R. Penndorf</i>	5
Observations of Direction of Arrival of Long-Duration Meteor Echoes in Forward Scatter Propagation,..... <i>T. Hagfors and B. Landmark</i>	19
Recurrent Geomagnetic Storms and Solar Prominences,..... <i>Richard T. Hansen</i>	23
An Analysis of Drifts of the Signal Pattern Associated with Ionospheric Reflections, <i>Donald G. Yerg</i>	27
Summer Upwelling Along the East Coast of Florida, <i>C. B. Taylor and H. B. Stewart, Jr.</i>	33
Tracing Beach Sand Movement with Irradiated Quartz, <i>D. L. Inman and T. K. Chamberlain</i>	41
Rapid Gravity Computations for Two-Dimensional Bodies with Application to the Mendocino Submarine Fracture Zone, <i>Manik Talwani, J. Lamar Worzel, and Mark Landisman</i>	49
Reconciliation of Stokes' Function and Astro-Geodetic Geoid Determinations, <i>W. M. Kaula</i>	61
A Tentative World Datum from Geoidal Heights Based on the Hough Ellipsoid and the Columbus Geoid,..... <i>Irene Fischer</i>	73
The Impact of the Ice Age on the Present Form of the Geoid,..... <i>Irene Fischer</i>	85
A Method of Evaluating the Effect of a Monomolecular Film in Suppressing Reservoir Evaporation,..... <i>G. Earl Harbeck, Jr., and Gordon E. Koberg</i>	89
A Note on the Propagation of a Sound Pulse in a Two-Layer Liquid Medium, <i>J. H. Rosenbaum</i>	95
Calculations Based on the Kozeny-Carman Theory,..... <i>Walter D. Rose</i>	103
Systematic Determination of Unit Hydrograph Parameters,..... <i>J. E. Nash</i>	111
On the Origin of Rock Magma,..... <i>Robert J. Uffen</i>	117
The Calorimetry of Steaming Ground in Thermal Areas,..... <i>R. F. Benseman</i>	123
The Origin of Some Natural Carbon Dioxide Gases,..... <i>Walter B. Lang</i>	127
Letter to the Editor: Discussion of "Frequency Curves for Peak Rates of Runoff," by W. D. Potter..... <i>G. N. Alexander</i>	132